

# **Zur Durchführung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung**

**Eine theoretische und empirische Analyse aus einer  
verhaltenswissenschaftlichen Perspektive**

**Vom Fachbereich genehmigte**

**Dissertation**

**zur Erlangung des akademischen Grades  
Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)  
des Fachbereiches Rechts- und Wirtschaftswissenschaften  
der Technischen Universität Darmstadt**

**vorgelegt von**

**Dipl.-Kfm. Engin Kayadelen**

**geboren in Gelsenkirchen**

**Erstreferent:  
Prof. Dr. Reiner Quick**

**Korreferent:  
Prof. Dr. Kai-Uwe Marten**

**Tag der Einreichung:  
12. Juni 2007**

**Tag der mündlichen Prüfung:  
13. Dezember 2007**

**D17**

**Darmstadt, 2008**

# Inhaltsverzeichnis

|                                    |             |
|------------------------------------|-------------|
| <b>ABBILDUNGSVERZEICHNIS .....</b> | <b>XII</b>  |
| <b>TABELLENVERZEICHNIS.....</b>    | <b>XIII</b> |
| <b>ABKÜRZUNGSVERZEICHNIS .....</b> | <b>XVI</b>  |
| <b>SYMBOLVERZEICHNIS .....</b>     | <b>XIX</b>  |

|                                  |          |
|----------------------------------|----------|
| <b>1 EINLEITENDER TEIL .....</b> | <b>1</b> |
| 1.1 Problemstellung .....        | 1        |
| 1.2 Gang der Untersuchung .....  | 6        |

|                                                                                             |           |
|---------------------------------------------------------------------------------------------|-----------|
| <b>2 ANALYTISCHE PRÜFUNGSHANDLUNGEN ALS GEGENSTAND DER<br/>JAHRESABSCHLUSSPRÜFUNG .....</b> | <b>10</b> |
|---------------------------------------------------------------------------------------------|-----------|

|                                                                                                                                        |           |
|----------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>2.1 Grundlegende Überlegungen.....</b>                                                                                              | <b>10</b> |
| 2.1.1 Begriffsbestimmung und Wesen analytischer Prüfungshandlungen .....                                                               | 10        |
| 2.1.2 Anwendungsgebiete und Ziele analytischer Prüfungshandlungen .....                                                                | 16        |
| 2.1.2.1 Vorbemerkung .....                                                                                                             | 16        |
| 2.1.2.2 Analytische Prüfungshandlungen im Rahmen der Prüfungsplanung.....                                                              | 17        |
| 2.1.2.2.1 Ziele analytischer Prüfungshandlungen im Rahmen der Prüfungsplanung.....                                                     | 17        |
| 2.1.2.2.1.1 Vertiefung der Kenntnisse über die Geschäftstätigkeit des Mandanten .....                                                  | 17        |
| 2.1.2.2.1.2 Identifikation von kritischen Prüfungsgebieten .....                                                                       | 20        |
| 2.1.2.2.1.3 Planung der Art, des Umfangs und des zeitlichen Ablaufs von weiteren<br>Prüfungshandlungen .....                           | 21        |
| 2.1.2.2.2 Prüfungsgegenstände analytischer Prüfungshandlungen im Rahmen der<br>Prüfungsplanung .....                                   | 23        |
| 2.1.2.3 Analytische Prüfungshandlungen im Rahmen der Prüfungsdurchführung .....                                                        | 24        |
| 2.1.2.3.1 Ziele analytischer Prüfungshandlungen im Rahmen der Prüfungsdurchführung und<br>Entscheidungskriterien ihres Einsatzes ..... | 24        |
| 2.1.2.3.1.1 Verlässlichkeit der Ergebnisse analytischer Prüfungshandlungen.....                                                        | 25        |
| 2.1.2.3.1.2 Art des Unternehmens und die Möglichkeit, Informationen wieder in ihre<br>Bestandteile zu zerlegen .....                   | 28        |
| 2.1.2.3.1.3 Verfügbarkeit der Informationen .....                                                                                      | 28        |

|               |                                                                                                           |           |
|---------------|-----------------------------------------------------------------------------------------------------------|-----------|
| 2.1.2.3.1.4   | Zuverlässigkeit der Informationen .....                                                                   | 29        |
| 2.1.2.3.1.5   | Relevanz der verfügbaren Informationen .....                                                              | 30        |
| 2.1.2.3.1.6   | Vergleichbarkeit der verfügbaren Informationen .....                                                      | 30        |
| 2.1.2.3.1.7   | Erkenntnisse aus der Prüfungsplanung, den Vorjahresprüfungen und der<br>Systemprüfung.....                | 31        |
| 2.1.2.3.2     | Arten analytischer Prüfungshandlungen im Rahmen der Prüfungsdurchführung...                               | 31        |
| 2.1.2.3.3     | Prüfungsgegenstände analytischer Prüfungshandlungen im Rahmen der<br>Prüfungsdurchführung.....            | 33        |
| 2.1.2.4       | Analytische Prüfungshandlungen im Rahmen der abschließenden Gesamtdurchsicht<br>.....                     | 34        |
| 2.1.2.4.1     | Ziele analytischer Prüfungshandlungen in der abschließenden Gesamtdurchsicht<br>.....                     | 34        |
| 2.1.2.4.1.1   | Vorbemerkungen .....                                                                                      | 34        |
| 2.1.2.4.1.2   | Berücksichtigung von Korrekturbuchungen oder Reklassifizierungen im<br>Jahresabschluss .....              | 35        |
| 2.1.2.4.1.3   | Nachträgliche Identifikation bisher nicht erkannter Fehlerrisiken.....                                    | 35        |
| 2.1.2.4.1.4   | Beurteilung der Angemessenheit der erlangten Prüfungsnachweise .....                                      | 36        |
| 2.1.2.4.2     | Prüfungsgegenstände analytischer Prüfungshandlungen in Rahmen der<br>abschließenden Gesamtdurchsicht..... | 37        |
| <b>2.2</b>    | <b>Verfahren analytischer Prüfungshandlungen .....</b>                                                    | <b>38</b> |
| 2.2.1         | Grundlegende Vorbemerkungen.....                                                                          | 38        |
| 2.2.2         | Informationen aus der Vergangenheit.....                                                                  | 39        |
| 2.2.2.1       | Einfacher Vorjahresvergleich .....                                                                        | 39        |
| 2.2.2.2       | Einfache Trendanalysen.....                                                                               | 41        |
| 2.2.2.2.1     | Vorbemerkungen.....                                                                                       | 41        |
| 2.2.2.2.2     | Vorgehensweise des Prüfers bei Trendanalysen .....                                                        | 42        |
| 2.2.2.2.3     | Methoden der Trendanalyse .....                                                                           | 43        |
| 2.2.2.2.3.1   | Übersicht.....                                                                                            | 43        |
| 2.2.2.2.3.2   | Graphische Methode .....                                                                                  | 43        |
| 2.2.2.2.3.3   | Methode der durchschnittlichen Änderung.....                                                              | 45        |
| 2.2.2.2.3.4   | Methode der gewichteten durchschnittlichen Änderung.....                                                  | 46        |
| 2.2.2.2.3.5   | Methode des gleitenden Durchschnitts .....                                                                | 46        |
| 2.2.2.2.3.6   | Methode der kleinsten Quadrate .....                                                                      | 47        |
| 2.2.2.3       | Zeitbezogene Kennzahlenanalysen .....                                                                     | 49        |
| 2.2.2.3.1     | Vorüberlegungen.....                                                                                      | 49        |
| 2.2.2.3.2     | Begriffsbestimmung.....                                                                                   | 49        |
| 2.2.2.3.3     | Arten von zeitbezogenen Kennzahlenanalysen .....                                                          | 51        |
| 2.2.2.3.4     | Geeignete Kennzahlen zur Beurteilung der Plausibilität von Jahresabschlussdaten<br>.....                  | 52        |
| 2.2.2.3.4.1   | Überblick.....                                                                                            | 52        |
| 2.2.2.3.4.2   | Bilanz .....                                                                                              | 53        |
| 2.2.2.3.4.2.1 | Sachanlagevermögen.....                                                                                   | 53        |

|               |                                                                                                            |           |
|---------------|------------------------------------------------------------------------------------------------------------|-----------|
| 2.2.2.3.4.2.2 | Vorräte .....                                                                                              | 55        |
| 2.2.2.3.4.2.3 | Forderungen .....                                                                                          | 57        |
| 2.2.2.3.4.2.4 | Verbindlichkeiten und Rückstellungen .....                                                                 | 59        |
| 2.2.2.3.4.3   | Gewinn- und Verlustrechnung .....                                                                          | 61        |
| 2.2.2.3.4.3.1 | Beurteilung der Umsatzerlöse .....                                                                         | 61        |
| 2.2.2.3.4.3.2 | Beurteilung der Erträge aus Zinsen und Dividenden .....                                                    | 62        |
| 2.2.2.3.4.3.3 | Beurteilung der Aufwendungen .....                                                                         | 63        |
| 2.2.2.3.5     | Probleme/Grenzen der Kennzahlenanalyse .....                                                               | 65        |
| 2.2.3         | Informationen aus dem aktuellen Geschäftsjahr .....                                                        | 68        |
| 2.2.3.1       | Benchmark-Vergleiche .....                                                                                 | 68        |
| 2.2.3.1.1     | Übersicht über die Arten von Benchmark-Vergleichen .....                                                   | 68        |
| 2.2.3.1.2     | Planzahlen-Benchmarking .....                                                                              | 68        |
| 2.2.3.1.3     | Branchen-Benchmarking .....                                                                                | 70        |
| 2.2.3.1.3.1   | Wesen des Branchen-Benchmarkings .....                                                                     | 70        |
| 2.2.3.1.3.2   | Arten des Branchen-Benchmarkings .....                                                                     | 71        |
| 2.2.3.1.3.2.1 | Branchendurchschnittswert-Benchmarking .....                                                               | 71        |
| 2.2.3.1.3.2.2 | Unternehmens-Benchmarking .....                                                                            | 73        |
| 2.2.3.1.3.3   | Vorteile und Nachteile des Branchen-Benchmarkings .....                                                    | 74        |
| 2.2.3.1.4     | Betriebs-Benchmarking .....                                                                                | 75        |
| 2.2.3.2       | Einfache Rechenmodelle .....                                                                               | 75        |
| 2.2.3.2.1     | Wesen einfacher Rechenmodelle .....                                                                        | 75        |
| 2.2.3.2.2     | Anwendungsbereiche von einfachen Rechenmodellen .....                                                      | 76        |
| 2.2.3.2.3     | Vorteile und Nachteile von Rechenmodellen .....                                                            | 77        |
| 2.2.4         | Informationen aus der Vergangenheit und aus dem aktuellen Geschäftsjahr .....                              | 81        |
| 2.2.4.1       | Einordnung mathematisch-statistischer Verfahren .....                                                      | 81        |
| 2.2.4.2       | Die Regressionsanalyse als beispielhafte Darstellung eines mathematisch-statistischen Verfahrens .....     | 81        |
| 2.2.4.2.1     | Begriffsbestimmung und Wesen von Regressionsanalysen .....                                                 | 81        |
| 2.2.4.2.2     | Formen von Regressionsanalysen .....                                                                       | 82        |
| 2.2.4.2.3     | Vorteile und Nachteile der Regressionsanalyse .....                                                        | 88        |
| <b>2.3</b>    | <b>Ablaufmodell bei der Durchführung analytischer Prüfungshandlungen .....</b>                             | <b>91</b> |
| 2.3.1         | Die Zerlegung der prüferischen Vorgehensweise in einzelne Prüfungsschritte .....                           | 91        |
| 2.3.2         | Beschreibung der einzelnen Prüfungsschritte bei der Durchführung von analytischen Prüfungshandlungen ..... | 92        |
| 2.3.2.1       | Bildung eines Erwartungswerts .....                                                                        | 92        |
| 2.3.2.1.1     | Einflussfaktoren der Qualität des Erwartungswerts .....                                                    | 92        |
| 2.3.2.1.2     | Verständnis des Prüfers für die Zusammenhänge in den jahresabschlussrelevanten Daten .....                 | 93        |
| 2.3.2.1.3     | Genauigkeit des entwickelten Erwartungswerts .....                                                         | 94        |
| 2.3.2.2       | Festlegung von Toleranzgrenzen .....                                                                       | 95        |
| 2.3.2.2.1     | Zweck und Zeitpunkt der Festlegung von Toleranzgrenzen .....                                               | 95        |
| 2.3.2.2.2     | Kriterien zur Festlegung von Toleranzgrenzen .....                                                         | 96        |

|            |                                                                                         |            |
|------------|-----------------------------------------------------------------------------------------|------------|
| 2.3.2.3    | Soll-Ist-Vergleich .....                                                                | 98         |
| 2.3.2.4    | Beurteilung .....                                                                       | 99         |
| 2.3.2.4.1  | Allgemeine Vorgehensweise im Rahmen der Beurteilung einer signifikanten Abweichung..... | 99         |
| 2.3.2.4.2  | Hypothesengenerierung .....                                                             | 100        |
| 2.3.2.4.3  | Informationsbeschaffung .....                                                           | 101        |
| 2.3.2.4.4  | Hypothesenbewertung .....                                                               | 102        |
| 2.3.2.4.5  | Prüfungsentscheidung.....                                                               | 103        |
| <b>2.4</b> | <b>Zwischenfazit .....</b>                                                              | <b>105</b> |

### **3 ANALYTISCHE PRÜFUNGSHANDLUNGEN ALS GEGENSTAND DER EMPIRISCHEN PRÜFUNGSFORSCHUNG..... 106**

|             |                                                                                                               |            |
|-------------|---------------------------------------------------------------------------------------------------------------|------------|
| <b>3.1</b>  | <b>Theoretische und konzeptionelle Vorbemerkungen .....</b>                                                   | <b>106</b> |
| 3.1.1       | Prüfungstheoretische Grundlagen .....                                                                         | 106        |
| 3.1.2       | Forschungsziel .....                                                                                          | 108        |
| 3.1.3       | Forschungsprogramme .....                                                                                     | 111        |
| <b>3.2</b>  | <b>Weiterführende Überlegungen.....</b>                                                                       | <b>112</b> |
| 3.2.1       | Einordnung des (Forschungs-)Objekts „Prüfungswesen“ in die theoretischen und konzeptionellen Grundlagen ..... | 112        |
| 3.2.1.1     | Prüfungswesen als realwissenschaftliche und angewandte Wissenschaft zugleich .                                | 112        |
| 3.2.1.2     | Ausgestaltung eines prüfungstheoretischen Forschungsprogramms .....                                           | 115        |
| 3.2.1.2.1   | Vorüberlegungen.....                                                                                          | 115        |
| 3.2.1.2.1.1 | Formal-analytische Forschungsprogramme.....                                                                   | 115        |
| 3.2.1.2.1.2 | Erfahrungswissenschaftliche Forschungsprogramme.....                                                          | 118        |
| 3.2.1.2.1.3 | Würdigung der bisherigen Überlegungen.....                                                                    | 123        |
| 3.2.1.2.2   | Skizzierung eines erfahrungswissenschaftlichen Forschungsprogramms für das Prüfungswesen.....                 | 125        |
| 3.2.2       | Verhaltensorientierte Modelle zur Beschreibung des prüferischen Vorgehens.....                                | 128        |
| 3.2.2.1     | Regressions- und Attributionsmodelle.....                                                                     | 128        |
| 3.2.2.2     | Modelle probabilistischer Wahrscheinlichkeitsschlüsse .....                                                   | 129        |
| 3.2.2.2.1   | Wesen probabilistischer Modelle .....                                                                         | 129        |
| 3.2.2.2.2   | Heuristiken bei Wahrscheinlichkeitsurteilen.....                                                              | 130        |
| 3.2.2.2.2.1 | Begrenzte Rationalität und Heuristiken .....                                                                  | 130        |
| 3.2.2.2.2.2 | Arten von Urteilsbildungsheuristiken .....                                                                    | 131        |
| 3.2.2.2.2.3 | Würdigung der Urteilsbildungsheuristiken.....                                                                 | 134        |
| 3.2.2.3     | Problemlösungsmodelle .....                                                                                   | 136        |
| 3.2.2.3.1   | Modelltheoretische Vorüberlegungen .....                                                                      | 136        |

|                 |                                                                                                                                                                          |            |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.2.2.3.1.1     | Der kognitionspsychologische Informationsverarbeitungsansatz als Ausgangspunkt der Betrachtung .....                                                                     | 136        |
| 3.2.2.3.1.2     | Skizzierung eines Problemlösungsmodells für die Jahresabschlussprüfung .....                                                                                             | 137        |
| 3.2.2.3.2       | Problembegriff .....                                                                                                                                                     | 139        |
| 3.2.2.3.3       | Problemlösungsstrategien .....                                                                                                                                           | 141        |
| 3.2.2.3.3.1     | Genereller Ablauf einer Problemlösung .....                                                                                                                              | 141        |
| 3.2.2.3.3.2     | Arten von Problemlösungsstrategien .....                                                                                                                                 | 142        |
| 3.2.2.3.3.2.1   | Versuch-Irrtum-Analyse .....                                                                                                                                             | 142        |
| 3.2.2.3.3.2.2   | Rückwärtsgerichtete Suchstrategien .....                                                                                                                                 | 142        |
| 3.2.2.3.3.2.2.1 | Mittel-Ziel-Analyse .....                                                                                                                                                | 142        |
| 3.2.2.3.3.2.2.2 | Generate and test-Strategie .....                                                                                                                                        | 144        |
| 3.2.2.3.3.2.3   | Vorwärtsgerichtete Suchstrategie .....                                                                                                                                   | 144        |
| 3.2.2.3.3.2.4   | Situations- und Zielanalyse .....                                                                                                                                        | 146        |
| 3.2.2.3.3.2.5   | Umstrukturierung und Neuentdeckung .....                                                                                                                                 | 147        |
| 3.2.2.3.3.2.5.1 | Einstellungs- und Fixierungseffekte .....                                                                                                                                | 147        |
| 3.2.2.3.3.2.5.2 | Umstrukturierungsheurismen .....                                                                                                                                         | 149        |
| 3.2.2.3.3.2.5.3 | Neuentdeckungsheurismen .....                                                                                                                                            | 150        |
| 3.2.2.3.4       | Wissen und Wissensstrukturen .....                                                                                                                                       | 152        |
| 3.2.2.3.4.1     | Der Wissensbegriff und die unterschiedlichen Arten des Wissens .....                                                                                                     | 152        |
| 3.2.2.3.4.2     | Wissensstrukturen .....                                                                                                                                                  | 154        |
| 3.2.2.3.4.2.1   | Eigenschaften von Wissensstrukturen .....                                                                                                                                | 154        |
| 3.2.2.3.4.2.2   | Typologien von Wissensstrukturen .....                                                                                                                                   | 155        |
| 3.2.2.3.4.2.2.1 | Überblick .....                                                                                                                                                          | 155        |
| 3.2.2.3.4.2.2.2 | Propositionale Modelle .....                                                                                                                                             | 156        |
| 3.2.2.3.4.2.2.3 | Regelbasierte Modelle .....                                                                                                                                              | 159        |
| 3.2.2.3.4.2.2.4 | Schemabasierte Modelle .....                                                                                                                                             | 161        |
| 3.2.2.4         | Auswahl eines verhaltensorientierten Modells zur Abbildung der prüferischen Vorgehensweise .....                                                                         | 164        |
| <b>3.3</b>      | <b>Entwicklung eines Ablaufmodells zur Durchführung von analytischen Prüfungshandlungen aus dem Blickwinkel eines realwissenschaftlichen Theorieverständnisses .....</b> | <b>166</b> |
| <b>3.4</b>      | <b>Empirische Forschungsmethoden: Methoden der Datengewinnung und -auswertung .....</b>                                                                                  | <b>172</b> |
| 3.4.1           | Vorbemerkungen .....                                                                                                                                                     | 172        |
| 3.4.2           | Methoden der Datengewinnung .....                                                                                                                                        | 172        |
| 3.4.2.1         | Befragung .....                                                                                                                                                          | 172        |
| 3.4.2.2         | Beobachtung .....                                                                                                                                                        | 174        |
| 3.4.2.3         | Inhaltsanalyse .....                                                                                                                                                     | 176        |
| 3.4.3           | Methoden der Datenauswertung .....                                                                                                                                       | 178        |

|                 |                                                                                                                                                                                       |            |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>3.5</b>      | <b>Darstellung und Analyse der empirischen Prüfungsforschung zur Durchführung von analytischen Prüfungshandlungen vor dem Hintergrund des abgeleiteten Problemlösungsmodells.....</b> | <b>180</b> |
| 3.5.1           | Vorbemerkung.....                                                                                                                                                                     | 180        |
| 3.5.2           | Studien zur Phase der Problemrepräsentation.....                                                                                                                                      | 180        |
| 3.5.2.1         | Aufbau einer adäquaten Problemrepräsentation und der Einfluss von Wissensstrukturen auf das Problemverständnis .....                                                                  | 180        |
| 3.5.2.2         | Verzerrungen beim Aufbau der Problemrepräsentation.....                                                                                                                               | 187        |
| 3.5.2.2.1       | Framing-Effekte.....                                                                                                                                                                  | 187        |
| 3.5.2.2.2       | Anker-Effekt (Anchoring) .....                                                                                                                                                        | 191        |
| 3.5.2.2.2.1     | Die Berücksichtigung ungeprüfter Buchwerte im Rahmen der Erwartungsbildung .....                                                                                                      | 191        |
| 3.5.2.2.2.2     | Debiasing als Möglichkeit zur Vermeidung des Anker-Effekts.....                                                                                                                       | 200        |
| 3.5.2.2.3       | Favourable Outcome-Effekt .....                                                                                                                                                       | 208        |
| 3.5.2.3         | Der Aufbau einer inadäquaten Problemrepräsentation durch den Erhalt einer Mandantenerklärung und der Wechsel zu einer adäquaten Problemrepräsentation.....                            | 211        |
| 3.5.2.4         | Tabellarische Darstellung der empirischen Befunde zur Phase der Problemrepräsentation .....                                                                                           | 214        |
| 3.5.3           | Studien zur Phase der Hypothesengenerierung.....                                                                                                                                      | 218        |
| 3.5.3.1         | Vorbemerkungen.....                                                                                                                                                                   | 218        |
| 3.5.3.2         | Unterteilung nach dem Zeitpunkt/-raum der Hypothesengenerierung .....                                                                                                                 | 219        |
| 3.5.3.3         | Unterteilung nach der Quelle der Hypothese .....                                                                                                                                      | 221        |
| 3.5.3.3.1       | Interne Hypothesengenerierung.....                                                                                                                                                    | 221        |
| 3.5.3.3.1.1     | Einflussfaktoren der ersten Hypothesengenerierung .....                                                                                                                               | 221        |
| 3.5.3.3.1.1.1   | Frequency knowledge .....                                                                                                                                                             | 221        |
| 3.5.3.3.1.1.2   | Recency effects .....                                                                                                                                                                 | 224        |
| 3.5.3.3.1.2     | Einfluss der ersten Hypothese auf die folgende Hypothesengenerierung ....                                                                                                             | 225        |
| 3.5.3.3.1.2.1   | Spreading effect .....                                                                                                                                                                | 225        |
| 3.5.3.3.1.2.2   | Interference effect.....                                                                                                                                                              | 227        |
| 3.5.3.3.1.3     | Anzahl der Hypothesen .....                                                                                                                                                           | 228        |
| 3.5.3.3.1.3.1   | Einfluss der Erfahrung auf die Anzahl der generierten plausiblen Hypothesen .....                                                                                                     | 228        |
| 3.5.3.3.1.3.2   | Einfluss der Anzahl der generierten Hypothesen auf die Effektivität und Effizienz von analytischen Prüfungshandlungen.....                                                            | 231        |
| 3.5.3.3.1.4     | Hypothesenwechsel .....                                                                                                                                                               | 233        |
| 3.5.3.3.2       | Externe Hypothesengenerierung .....                                                                                                                                                   | 235        |
| 3.5.3.3.2.1     | Der Einfluss des Erhalts einer Hypothese durch den Mandanten bzw. dessen Mitarbeiter.....                                                                                             | 235        |
| 3.5.3.3.2.1.1   | Vorbemerkungen .....                                                                                                                                                                  | 235        |
| 3.5.3.3.2.1.2   | Einfluss auf die nachfolgende interne Hypothesengenerierung.....                                                                                                                      | 236        |
| 3.5.3.3.2.1.2.1 | Interference effect.....                                                                                                                                                              | 236        |
| 3.5.3.3.2.1.2.2 | spreading effect .....                                                                                                                                                                | 238        |

|                 |                                                                                                                                                                          |     |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.5.3.3.2.1.2.3 | Bewertung der Forschungsergebnisse zum interference effect und spreading effect .....                                                                                    | 238 |
| 3.5.3.3.2.1.3   | Der Einfluss des Erhalts einer Mandantenerklärung auf den Erfolg einer analytischen Prüfungshandlung .....                                                               | 239 |
| 3.5.3.3.2.1.3.1 | Vorbemerkungen .....                                                                                                                                                     | 239 |
| 3.5.3.3.2.1.3.2 | Untersuchungsdesign ohne Kontrollgruppe .....                                                                                                                            | 239 |
| 3.5.3.3.2.1.3.3 | Untersuchung mit Kontrollgruppe .....                                                                                                                                    | 240 |
| 3.5.3.3.2.1.3.4 | Bewertung .....                                                                                                                                                          | 242 |
| 3.5.3.3.2.1.4   | Der Einfluss der Prüfererfahrung auf den Erfolg analytischer Prüfungshandlungen bei Erhalt einer Mandantenerklärung .....                                                | 242 |
| 3.5.3.3.2.1.4.1 | Einfluss von allgemeiner und domänenspezifischer Erfahrung .....                                                                                                         | 242 |
| 3.5.3.3.2.1.4.2 | Einfluss von branchenspezifischer Erfahrung .....                                                                                                                        | 244 |
| 3.5.3.3.2.2     | Der Einfluss des Erhalts einer Hypothese durch den Vorgesetzten auf die nachfolgende interne Hypothesengenerierung .....                                                 | 245 |
| 3.5.3.3.2.3     | Der Einfluss des Erhalts einer Hypothese durch Entscheidungshilfen .....                                                                                                 | 248 |
| 3.5.3.3.2.3.1   | Vorbemerkungen .....                                                                                                                                                     | 248 |
| 3.5.3.3.2.3.2   | Der Einfluss der Auswahlstrategie auf die Anzahl der ausgewählten Hypothesen .....                                                                                       | 249 |
| 3.5.3.4         | Der Einfluss von Gruppenprozessen auf die Hypothesengenerierung .....                                                                                                    | 251 |
| 3.5.3.4.1       | Vorbemerkungen .....                                                                                                                                                     | 251 |
| 3.5.3.4.2       | Der Einfluss des Review-Prozesses auf die Anzahl der plausiblen Hypothesen .....                                                                                         | 252 |
| 3.5.3.4.3       | Der Einfluss der Kommunikation unter den Mitgliedern eines Prüfungsteams auf die Anzahl der plausiblen Hypothesen und den Erfolg der analytischen Prüfungshandlung ..... | 254 |
| 3.5.3.5         | Tabellarische Darstellung der empirischen Befunde zur Phase der Hypothesengenerierung .....                                                                              | 257 |
| 3.5.4           | Studien zur Phase der Informationssuche .....                                                                                                                            | 265 |
| 3.5.4.1         | Vorbemerkungen .....                                                                                                                                                     | 265 |
| 3.5.4.2         | Verzerrungen im Rahmen der Informationssuche .....                                                                                                                       | 266 |
| 3.5.4.2.1       | Konfirmatorische Prozesse und positive Suchstrategie .....                                                                                                               | 266 |
| 3.5.4.2.2       | Commitment als wesentlicher Einflussfaktor der Verzerrungen .....                                                                                                        | 269 |
| 3.5.4.3         | Eigenschaften der Informationssuche und ihr Einfluss auf den Erfolg analytischer Prüfungshandlungen .....                                                                | 272 |
| 3.5.4.4         | Einflussfaktoren der Informationssuche .....                                                                                                                             | 275 |
| 3.5.4.4.1       | Vorbemerkungen .....                                                                                                                                                     | 275 |
| 3.5.4.4.2       | Rechtfertigungsdruck und zeitliche Beschränkungen .....                                                                                                                  | 275 |
| 3.5.4.4.3       | Der Einfluss des Erhalts einer Mandantenerklärung auf die Eigenschaften der Informationssuche mit und ohne Berücksichtigung der branchenspezifischen Erfahrung .....     | 277 |
| 3.5.4.4.3.1     | Ohne Berücksichtigung der Branchenerfahrung .....                                                                                                                        | 277 |
| 3.5.4.4.3.2     | Mit Berücksichtigung der Branchenerfahrung .....                                                                                                                         | 278 |



|                   |                                                                                                                      |     |
|-------------------|----------------------------------------------------------------------------------------------------------------------|-----|
| 3.5.4.5           | Tabellarische Darstellung der empirischen Befunde zur Phase der Informationssuche .....                              | 280 |
| 3.5.5             | Studien zur Phase der Hypothesenbewertung .....                                                                      | 285 |
| 3.5.5.1           | Vorbemerkung .....                                                                                                   | 285 |
| 3.5.5.2           | Generelle Erwartungshaltung von Prüfern gegenüber festgestellten Abweichungen                                        | 285 |
| 3.5.5.3           | Die Bewertung einer „vererbten“ Hypothese .....                                                                      | 286 |
| 3.5.5.3.1         | Der Erhalt einer Abweichungshypothese durch den Mandanten bzw. dessen Mitarbeiter .....                              | 286 |
| 3.5.5.3.1.1       | Der Einfluss des Erhalts einer Mandantenerklärung auf die Hypothesenbewertung .....                                  | 286 |
| 3.5.5.3.1.2       | Verzerrungen im Rahmen der Hypothesenbewertung sowie potenzielle debiasing-Mechanismen .....                         | 288 |
| 3.5.5.3.1.2.1     | Der Einfluss der Dokumentation der Mandantenerklärung .....                                                          | 288 |
| 3.5.5.3.1.2.1.1   | Vorbemerkungen .....                                                                                                 | 288 |
| 3.5.5.3.1.2.1.2   | Einfluss einer teilweise korrekten Mandantenerklärung .....                                                          | 289 |
| 3.5.5.3.1.2.1.3   | Einfluss einer vollständig inkorrekten Mandantenerklärung .....                                                      | 291 |
| 3.5.5.3.1.2.1.4   | Mögliche Debiasing-Mechanismen gegen den written explanation effect .....                                            | 292 |
| 3.5.5.3.1.2.1.4.1 | Dokumentation von Gegenargumenten .....                                                                              | 292 |
| 3.5.5.3.1.2.1.4.2 | Quantifizierung der Mandantenerklärung .....                                                                         | 293 |
| 3.5.5.3.1.2.1.5   | Würdigung .....                                                                                                      | 294 |
| 3.5.5.3.1.2.2     | Der Einfluss der Verständlichkeit einer Mandantenerklärung .....                                                     | 295 |
| 3.5.5.3.1.2.3     | Der Einfluss der Kompetenz der Informationsquelle .....                                                              | 297 |
| 3.5.5.3.1.2.4     | Einfluss der Integrität bzw. Objektivität der Informationsquelle .....                                               | 299 |
| 3.5.5.3.1.3       | Der Einfluss des Erhalts einer Mandantenerklärung auf den (Gesamt-) Erfolg einer analytischen Prüfungshandlung ..... | 300 |
| 3.5.5.3.2         | Der Erhalt einer Abweichungshypothese durch Entscheidungshilfen .....                                                | 301 |
| 3.5.5.3.2.1       | Vorbemerkungen .....                                                                                                 | 301 |
| 3.5.5.3.2.2       | Der Einfluss einer Fehler- bzw. Nicht-Fehler-Dominanz von Entscheidungshilfen auf die Hypothesenbewertung .....      | 301 |
| 3.5.5.3.2.3       | Der Einfluss der Verlässlichkeitseinschätzung auf die Hypothesenbewertung .....                                      | 303 |
| 3.5.5.4           | Die Bewertung einer Vielzahl von Hypothesen .....                                                                    | 304 |
| 3.5.5.4.1         | Vorbemerkung .....                                                                                                   | 304 |
| 3.5.5.4.2         | Verfahren zur Beurteilung einer Vielzahl von Hypothesen .....                                                        | 305 |
| 3.5.5.4.3         | Die Konjunktionsregel bei der Beurteilung mehrerer Hypothesen .....                                                  | 308 |
| 3.5.5.5           | Tabellarische Darstellung der empirischen Befunde zur Phase der Hypothesenbewertung .....                            | 311 |
| 3.5.6             | Studien zur Phase der Prüfungsentscheidung .....                                                                     | 317 |
| 3.5.6.1           | Vorbemerkung .....                                                                                                   | 317 |
| 3.5.6.2           | Die Prüfungsentscheidung im Rahmen des Einsatzes zur Prüfungsplanung .....                                           | 318 |
| 3.5.6.2.1         | Experimentelle Studien .....                                                                                         | 318 |

|             |                                                                                                                          |            |
|-------------|--------------------------------------------------------------------------------------------------------------------------|------------|
| 3.5.6.2.2   | Fragebogengestützte Studien .....                                                                                        | 320        |
| 3.5.6.2.3   | Der Einfluss des Erhalts einer Mandantenerklärung auf die Prüfungsentscheidung .....                                     | 321        |
| 3.5.6.3     | Tabellarische Darstellung der empirischen Befunde zur Phase der Prüfungsentscheidung.....                                | 325        |
| <b>3.6</b>  | <b>Zwischenbetrachtung .....</b>                                                                                         | <b>328</b> |
| <br>        |                                                                                                                          |            |
| <b>4</b>    | <b>EMPIRISCHE STUDIE ZUR DURCHFÜHRUNG VON ANALYTISCHEN PRÜFUNGSHANDLUNGEN IM RAHMEN DER JAHRESABSCHLUSSPRÜFUNG .....</b> | <b>330</b> |
| <br>        |                                                                                                                          |            |
| <b>4.1</b>  | <b>Vorbemerkungen.....</b>                                                                                               | <b>330</b> |
| <br>        |                                                                                                                          |            |
| <b>4.2</b>  | <b>Formulierung der Hypothesen und Forschungsfragen .....</b>                                                            | <b>332</b> |
| 4.2.1       | Der Einfluss des Erhalts einer Mandantenerklärung auf die interne Hypothesengenerierung .....                            | 332        |
| 4.2.2       | Der Einfluss des Erhalts einer Mandantenerklärung auf die Informationssuche .....                                        | 336        |
| 4.2.2.1     | Der Einfluss auf die angewandte Suchstrategie .....                                                                      | 336        |
| 4.2.2.2     | Der Einfluss auf die Interpretation der eingeholten Prüfungsnachweise .....                                              | 338        |
| 4.2.3       | Der Einfluss des Erhalts einer Mandantenerklärung auf die Hypothesenbewertung .....                                      | 340        |
| 4.2.4       | Der Einfluss des Erhalts einer Mandantenerklärung auf die Erfolgsquote .....                                             | 342        |
| <br>        |                                                                                                                          |            |
| <b>4.3</b>  | <b>Forschungsdesign .....</b>                                                                                            | <b>344</b> |
| 4.3.1       | Überblick über die ausgewählte Forschungsmethode und die Untersuchungsteilnehmer.....                                    | 344        |
| 4.3.2       | Beschreibung des Falls und des Ablaufs der Fallbearbeitung.....                                                          | 346        |
| <br>        |                                                                                                                          |            |
| <b>4.4</b>  | <b>Untersuchungsergebnisse.....</b>                                                                                      | <b>350</b> |
| 4.4.1       | Erläuterungen zu den durchgeführten Testverfahren .....                                                                  | 350        |
| 4.4.2       | Manipulationscheck.....                                                                                                  | 350        |
| 4.4.3       | Hypothesengenerierung.....                                                                                               | 351        |
| 4.4.3.1     | Vorbemerkungen .....                                                                                                     | 351        |
| 4.4.3.2     | Anfängliche Hypothesengenerierung .....                                                                                  | 352        |
| 4.4.3.2.1   | Test von $H_{l,1}$ und $H_{l,2}$ .....                                                                                   | 352        |
| 4.4.3.2.2   | Test von $H_{l,1,1}$ bis $H_{l,2,2}$ .....                                                                               | 353        |
| 4.4.3.2.3   | Test von $FF_{l,1}$ .....                                                                                                | 356        |
| 4.4.3.3     | Gesamte Hypothesengenerierung .....                                                                                      | 357        |
| 4.4.3.3.1   | Test von $H_{l,3}$ bis $H_{l,4}$ .....                                                                                   | 357        |
| 4.4.3.3.1.1 | Fehler-Version .....                                                                                                     | 357        |
| 4.4.3.3.1.2 | Nicht-Fehler-Version.....                                                                                                | 357        |

|             |                                                                  |     |
|-------------|------------------------------------------------------------------|-----|
| 4.4.3.3.2   | Test von $H_{I,3,1}$ bis $H_{I,4,2}$ .....                       | 358 |
| 4.4.3.3.2.1 | Fehler-Version .....                                             | 358 |
| 4.4.3.3.2.2 | Nicht-Fehler-Version.....                                        | 360 |
| 4.4.3.3.3   | Test von $FF_{I,2}$ .....                                        | 362 |
| 4.4.3.3.3.1 | Fehler-Version .....                                             | 362 |
| 4.4.3.3.3.2 | Nicht-Fehler-Version.....                                        | 362 |
| 4.4.4       | Informationssuche .....                                          | 363 |
| 4.4.4.1     | Vorbemerkungen .....                                             | 363 |
| 4.4.4.2     | Der Einfluss auf die angewandte Suchstrategie .....              | 364 |
| 4.4.4.2.1   | Test von $H_{II,1}$ und $H_{II,2}$ .....                         | 364 |
| 4.4.4.2.1.1 | Fehler-Version .....                                             | 364 |
| 4.4.4.2.1.2 | Nicht-Fehler-Version.....                                        | 365 |
| 4.4.4.2.2   | Test von $H_{II,1,1}$ bis $H_{II,2,2}$ .....                     | 366 |
| 4.4.4.2.2.1 | Fehler-Version .....                                             | 366 |
| 4.4.4.2.2.2 | Nicht-Fehler-Version.....                                        | 369 |
| 4.4.4.2.3   | Test von $FF_{II}$ .....                                         | 372 |
| 4.4.4.2.3.1 | Fehler-Version .....                                             | 372 |
| 4.4.4.2.3.2 | Nicht-Fehler-Version.....                                        | 375 |
| 4.4.4.3     | Der Einfluss auf die Interpretation von Prüfungsnachweisen ..... | 378 |
| 4.4.4.3.1   | Vorbemerkungen.....                                              | 378 |
| 4.4.4.3.2   | Test von $H_{III,1}$ und $H_{III,2}$ .....                       | 378 |
| 4.4.4.3.2.1 | Fehler-Version .....                                             | 378 |
| 4.4.4.3.2.2 | Nicht-Fehler-Version.....                                        | 379 |
| 4.4.4.3.3   | Test von $H_{III,1,1}$ bis $H_{III,2,2}$ .....                   | 380 |
| 4.4.4.3.3.1 | Fehler-Version .....                                             | 380 |
| 4.4.4.3.3.2 | Nicht-Fehler-Version.....                                        | 382 |
| 4.4.4.3.4   | Test von $FF_{III}$ .....                                        | 384 |
| 4.4.4.3.4.1 | Fehler-Version .....                                             | 384 |
| 4.4.4.3.4.2 | Nicht-Fehler-Version.....                                        | 385 |
| 4.4.5       | Hypothesenbewertung .....                                        | 386 |
| 4.4.5.1     | Test von $H_{IV,1,1}$ .....                                      | 386 |
| 4.4.5.1.1   | Fehler-Version.....                                              | 386 |
| 4.4.5.1.2   | Nicht-Fehler-Version .....                                       | 386 |
| 4.4.5.2     | Test von $H_{IV,1,1}$ und $H_{IV,1,2}$ .....                     | 387 |
| 4.4.5.2.1   | Fehler-Version.....                                              | 387 |
| 4.4.5.2.2   | Nicht-Fehler-Version .....                                       | 388 |
| 4.4.5.3     | Test von $FF_{IV}$ .....                                         | 389 |
| 4.4.5.3.1   | Fehler-Version.....                                              | 389 |
| 4.4.5.3.2   | Nicht-Fehler-Version .....                                       | 390 |
| 4.4.6       | Erfolgsquote .....                                               | 390 |
| 4.4.6.1     | Test von $H_{V,1}$ .....                                         | 390 |
| 4.4.6.1.1   | Fehler-Version.....                                              | 390 |

|            |                                                                    |            |
|------------|--------------------------------------------------------------------|------------|
| 4.4.6.1.2  | Nicht-Fehler-Version .....                                         | 391        |
| 4.4.6.2    | Test von $H_{V,1,1}$ und $H_{V,1,2}$ .....                         | 391        |
| 4.4.6.2.1  | Fehler-Version .....                                               | 391        |
| 4.4.6.2.2  | Nicht-Fehler-Version .....                                         | 392        |
| 4.4.6.3    | Test von $FF_V$ .....                                              | 393        |
| 4.4.6.3.1  | Fehler-Version .....                                               | 393        |
| 4.4.6.3.2  | Nicht-Fehler-Version .....                                         | 394        |
| <b>4.5</b> | <b>Zusammenfassung der Ergebnisse und Schlussfolgerungen .....</b> | <b>395</b> |
| 4.5.1      | Vorbemerkungen.....                                                | 395        |
| 4.5.2      | Anfängliche Hypothesengenerierung.....                             | 395        |
| 4.5.3      | Gesamte Hypothesengenerierung .....                                | 398        |
| 4.5.4      | Informationssuche .....                                            | 401        |
| 4.5.4.1    | Angewandte Suchstrategie .....                                     | 401        |
| 4.5.4.2    | Interpretation von Prüfungsnachweisen .....                        | 404        |
| 4.5.5      | Hypothesenbewertung .....                                          | 406        |
| 4.5.6      | Erfolgsquote .....                                                 | 408        |
| <b>4.6</b> | <b>Diskussion .....</b>                                            | <b>410</b> |
| <b>4.7</b> | <b>Grenzen der Untersuchung .....</b>                              | <b>414</b> |
| <b>5</b>   | <b>ZUSAMMENFASSUNG UND FAZIT .....</b>                             | <b>416</b> |
|            | <b>LITERATURVERZEICHNIS.....</b>                                   | <b>421</b> |
|            | <b>ANHANG 1: TEXT DER SAMMELMAIL .....</b>                         | <b>455</b> |
|            | <b>ANHANG 2: TEXT DES ANSCHREIBENS .....</b>                       | <b>456</b> |
|            | <b>ANHANG 3: LISTE DER 19 MÖGLICHEN TESTS .....</b>                | <b>458</b> |
|            | <b>ANHANG 4: GRUNDVERSION EINER FALLSTUDIE.....</b>                | <b>459</b> |

## **Abbildungsverzeichnis**

|                                                                                                                                                                    |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Abbildung 1: Ziele und Gegenstände analytischer Prüfungshandlungen in Abhängigkeit der Phase der Abschlussprüfung .....</b>                                     | <b>37</b>  |
| <b>Abbildung 2: Regressionsgerade zur Beurteilung der ausgewiesenen Forderungen aLuL anhand der Umsatzerlöse .....</b>                                             | <b>84</b>  |
| <b>Abbildung 3: Ablaufmodell zur Durchführung von analytischen Prüfungshandlungen .....</b>                                                                        | <b>92</b>  |
| <b>Abbildung 4: Erfahrungswissenschaftliches Forschungsprogramm für das Prüfungswesen .</b>                                                                        | <b>127</b> |
| <b>Abbildung 5: Darstellung der Problemelemente .....</b>                                                                                                          | <b>140</b> |
| <b>Abbildung 6: Taxonomie am Beispiel von ergebnisorientierten Prüfungshandlungen .....</b>                                                                        | <b>158</b> |
| <b>Abbildung 7: Ablaufmodell der Durchführung einer analytischen Prüfungshandlung aus dem Blickwinkel eines realwissenschaftlichen Theorieverständnisses .....</b> | <b>171</b> |

## Tabellenverzeichnis

|                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabelle 1: Ergebnisse der empirischen Prüfungsforschung zur Phase der Problemrepräsentation.....                                                    | 217 |
| Tabelle 2: Gegenüberstellung der durchschnittlichen Prüfungserfahrung und der durchschnittlichen Anzahl der generierten plausiblen Hypothesen ..... | 229 |
| Tabelle 3: Ergebnisse der empirischen Prüfungsforschung zur Phase der Hypothesengenerierung.....                                                    | 264 |
| Tabelle 4: Ergebnisse der empirischen Prüfungsforschung zur Phase der Informationssuche .....                                                       | 284 |
| Tabelle 5: Ergebnisse der empirischen Prüfungsforschung zur Phase der Hypothesenbewertung .....                                                     | 316 |
| Tabelle 6: Ergebnisse der empirischen Prüfungsforschung zur Phase der Prüfungsentscheidung .....                                                    | 327 |
| Tabelle 7: Umfang der durchschnittlichen Erfahrungsarten der Spezialisten und Nicht-Spezialisten .....                                              | 345 |
| Tabelle 8: Risikobeurteilung der Prüfergruppe mit niedrigem und erhöhtem Prüfungsrisiko ..                                                          | 351 |
| Tabelle 9: Anfängliche Hypothesengenerierung der Prüfer .....                                                                                       | 352 |
| Tabelle 10: Anfängliche Hypothesengenerierung der Nicht-Spezialisten .....                                                                          | 353 |
| Tabelle 11: Anfängliche Hypothesengenerierung der Spezialisten.....                                                                                 | 354 |
| Tabelle 12: Anfängliche Hypothesengenerierung der Prüfer unter Berücksichtigung des Prüfungsrisikos .....                                           | 356 |
| Tabelle 13: Gesamte Hypothesengenerierung der Prüfer (Fehler-Version).....                                                                          | 357 |
| Tabelle 14: Gesamte Hypothesengenerierung der Prüfer (Nicht-Fehler-Version) .....                                                                   | 358 |
| Tabelle 15: Gesamte Hypothesengenerierung der Nicht-Spezialisten (Fehler-Version) .....                                                             | 359 |
| Tabelle 16: Gesamte Hypothesengenerierung der Spezialisten (Fehler-Version) .....                                                                   | 359 |
| Tabelle 17: Gesamte Hypothesengenerierung der Nicht-Spezialisten (Nicht-Fehler-Version) ..                                                          | 360 |
| Tabelle 18: Gesamte Hypothesengenerierung der Spezialisten (Nicht-Fehler-Version) .....                                                             | 361 |
| Tabelle 19: Gesamte Hypothesengenerierung der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version).....                               | 362 |
| Tabelle 20: Gesamte Hypothesengenerierung der Prüfer mit Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version) .....                          | 363 |
| Tabelle 21: Angewandte Suchstrategie der Prüfer (Fehler-Version) .....                                                                              | 365 |
| Tabelle 22: Angewandte Suchstrategie der Prüfer (Nicht-Fehler-Version).....                                                                         | 366 |
| Tabelle 23: Breite der Suche und Tiefe der Kat. 1-Tests der Nicht-Spezialisten (Fehler-Version) .....                                               | 367 |
| Tabelle 24: Tiefe der Kat. 4 & 5-Tests der Nicht-Spezialisten (Fehler-Version).....                                                                 | 367 |
| Tabelle 25: Empfundenes Prüfungsrisiko der Nicht-Spezialisten in Abhängigkeit des Erhalts einer Mandantenerklärung.....                             | 368 |
| Tabelle 26: Breite der Suche und Tiefe der Kat. 1-Tests der Spezialisten (Fehler-Version) .....                                                     | 369 |
| Tabelle 27: Breite der Suche und Tiefe der Kat. 1-Tests der Nicht-Spezialisten (Nicht-Fehler-Version) .....                                         | 369 |

|                                                                                                                                                                                    |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabelle 28: Tiefe der Kat. 4 & 5-Tests der Nicht-Spezialisten (Nicht-Fehler-Version) .....                                                                                         | 370 |
| Tabelle 29: Breite der Suche und Tiefe der Kat. 1-Tests der Spezialisten (Nicht-Fehler-Version)<br>.....                                                                           | 371 |
| Tabelle 30: Umfang der Suche sowie Tiefe der Kat. 4- und Kat. 5-Tests der Spezialisten (Nicht-Fehler-Version) .....                                                                | 372 |
| Tabelle 31: Angewandte Suchstrategie der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version) .....                                                                  | 374 |
| Tabelle 32: Angewandte Suchstrategie der Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version) .....                                                            | 377 |
| Tabelle 33: : Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer (Fehler-Version) .....                                                        | 379 |
| Tabelle 34: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Prüfer (Fehler-Version) .....                                            | 379 |
| Tabelle 35: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer (Nicht-Fehler-Version) .....                                                    | 380 |
| Tabelle 36: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Nicht-Spezialisten (Fehler-Version) .....                                              | 380 |
| Tabelle 37: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Nicht-Spezialisten (Fehler-Version) .....                                | 381 |
| Tabelle 38: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Spezialisten (Fehler-Version) .....                                                    | 382 |
| Tabelle 39: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Spezialisten (Fehler-Version) .....                                      | 382 |
| Tabelle 40: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Nicht-Spezialisten (Nicht-Fehler-Version) .....                                        | 383 |
| Tabelle 41: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Spezialisten (Nicht-Fehler-Version) .....                                              | 383 |
| Tabelle 42: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version) ..... | 384 |
| Tabelle 43: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version) .....               | 385 |
| Tabelle 44: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version) .....         | 386 |
| Tabelle 45: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer (Fehler-Version) .....                                                                                           | 386 |
| Tabelle 46: Auswahlhäufigkeit der Mandantenerklärung ohne Berücksichtigung der branchenspezifischen Erfahrung (Nicht-Fehler-Version) .....                                         | 387 |
| Tabelle 47: Auswahlhäufigkeit der Mandantenerklärung durch Nicht-Spezialisten (Fehler-Version) .....                                                                               | 387 |
| Tabelle 48: Auswahlhäufigkeit der Mandantenerklärung durch Spezialisten (Fehler-Version) .....                                                                                     | 388 |
| Tabelle 49: Auswahlhäufigkeit der Mandantenerklärung durch Nicht-Spezialisten (Nicht-Fehler-Version) .....                                                                         | 388 |

|                                                                                                                                                  |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Tabelle 50: Auswahlhäufigkeit der Mandantenerklärung durch Spezialisten (Nicht-Fehler-Version) .....</b>                                      | <b>389</b> |
| <b>Tabelle 51: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version) .....</b>       | <b>389</b> |
| <b>Tabelle 52: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version) .....</b> | <b>390</b> |
| <b>Tabelle 53: Erfolgsquote der Prüfer (Fehler-Version) .....</b>                                                                                | <b>391</b> |
| <b>Tabelle 54: Erfolgsquote der Prüfer (Nicht-Fehler-Version) .....</b>                                                                          | <b>391</b> |
| <b>Tabelle 55: Erfolgsquote der Nicht-Spezialisten (Fehler-Version) .....</b>                                                                    | <b>392</b> |
| <b>Tabelle 56: Erfolgsquote der Nicht-Spezialisten (Fehler-Version) .....</b>                                                                    | <b>392</b> |
| <b>Tabelle 57: Erfolgsquote der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version) .....</b>                                     | <b>393</b> |
| <b>Tabelle 58: Erfolgsquote der Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version) .....</b>                               | <b>394</b> |
| <b>Tabelle 59: Zusammenfassung der Untersuchungsergebnisse (anfängliche Hypothesengenerierung) .....</b>                                         | <b>397</b> |
| <b>Tabelle 60: Zusammenfassung der Untersuchungsergebnisse (gesamte Hypothesengenerierung) .....</b>                                             | <b>400</b> |
| <b>Tabelle 61: Zusammenfassung der Untersuchungsergebnisse (angewandte Suchstrategie) ..</b>                                                     | <b>403</b> |
| <b>Tabelle 62: Zusammenfassung der Untersuchungsergebnisse (Interpretation der Prüfungsergebnisse) .....</b>                                     | <b>405</b> |
| <b>Tabelle 63: Zusammenfassung der Untersuchungsergebnisse (Hypothesenbewertung) .....</b>                                                       | <b>407</b> |
| <b>Tabelle 64: Zusammenfassung der Untersuchungsergebnisse (Erfolgsquote) .....</b>                                                              | <b>409</b> |



## Abkürzungsverzeichnis

|         |                                   |
|---------|-----------------------------------|
| Abb.    | Abbildung                         |
| Abs.    | Absatz                            |
| ACT     | Adaptive Control of Thoughts      |
| aLuL    | aus Lieferungen und Leistungen    |
| AP      | Abschlussprüfer                   |
| Aufl.   | Auflage                           |
| bspw.   | beispielsweise                    |
| BW      | Buchwert                          |
| bzgl.   | bezüglich                         |
| bzw.    | beziehungsweise                   |
| d.h.    | das heißt                         |
| EDV     | Elektronische Datenverarbeitung   |
| et al.  | und Andere                        |
| EW      | Erwartungswert                    |
| f.      | folgende                          |
| FF      | Forschungsfrage                   |
| Fn.     | Fußnote                           |
| ggü.    | gegenüber                         |
| griech. | griechisch                        |
| H       | Hypothese                         |
| HGB     | Handelsgesetzbuch                 |
| Hrsg.   | Herausgeber                       |
| http    | hypertext transfer protocol       |
| i.d.R.  | in der Regel                      |
| i.H.v.  | in Höhe von                       |
| i.V.m.  | in Verbindung mit                 |
| IAS     | International Accounting Standard |

|        |                                        |
|--------|----------------------------------------|
| IDW    | Institut der Wirtschaftsprüfer         |
| ISA    | International Standard on Auditing     |
| IT     | Informationstechnologie                |
| J.     | Jahr (-e)                              |
| JA     | Jahresabschluss                        |
| k.A.   | keine Angabe                           |
| Kap.   | Kapitel                                |
| Kat.   | Kategorie                              |
| Km     | Kilometer                              |
| LKW    | Lastkraftwagen                         |
| m.w.N. | mit weiteren Nachweisen                |
| min.   | Minute (-n)                            |
| Mio.   | Million (-en)                          |
| Mon.   | Monat (-e)                             |
| n.a.   | nicht anwendbar                        |
| Nr.    | Nummer                                 |
| o.a.   | oben aufgeführt (-e)/(-en)/(-em)/(-es) |
| o.Ä.   | oder Ähnliches                         |
| OG     | Obergrenze                             |
| p      | Signifikanzniveau                      |
| PS     | Prüfungsstandard                       |
| S.     | Seite (-n)                             |
| Sign.  | Signifikanz                            |
| sog.   | sogenannte (-s)/(-r)/(-n)              |
| Sp.    | Spalte (-n)                            |
| Std.   | Stunde (-n)                            |
| Tab.   | Tabelle                                |
| u.a.   | unter anderem                          |

|      |                         |
|------|-------------------------|
| u.U. | unter Umständen         |
| UG   | Untergrenze             |
| usw. | und so weiter           |
| Vgl. | Vergleich (-e)          |
| vs.  | versus                  |
| WPK  | Wirtschaftsprüferkammer |
| www  | world wide web          |
| z.B. | zum Beispiel            |
| z.T. | zum Teil                |

## Symbolverzeichnis

|           |                      |
|-----------|----------------------|
| \$        | US-Dollar            |
| %         | Prozent              |
| $\bar{x}$ | Durchschnitt (-lich) |
| $\mu$     | Mittelwert           |
| $\sigma$  | Standardabweichung   |
| &         | und                  |
| §         | Paragraf             |
| €         | Euro                 |

# 1 Einleitender Teil

## 1.1 Problemstellung

Ein Formalziel der Jahresabschlussprüfung ist die Erlangung eines vertrauenswürdigen Urteils über die Normenkonformität des Jahresabschlusses und des Lageberichts.<sup>1</sup> Hierzu obliegt es dem Prüfer, Prüfungsnachweise einzuholen, welche die Existenz oder Nicht-Existenz von wesentlichen Fehlern im Jahresabschluss und Lagebericht dokumentieren. Neben der Anzahl der gewonnenen Prüfungsnachweise – und somit der Menge der durchgeführten Prüfungshandlungen – ist insbesondere auch die *Wirksamkeit* der eingesetzten Prüfungshandlungen von zentraler Bedeutung für die Vertrauenswürdigkeit des Prüfungsurteils. Unter Wirksamkeit (im Folgenden auch als *Effektivität* bezeichnet) ist dabei die Fähigkeit zur Aufdeckung von wesentlichen Fehlern im Jahresabschluss und Lagebericht zu verstehen. Gleichzeitig stellt jedoch auch die *Wirtschaftlichkeit* der Urteilsgewinnung ein Formalziel der Jahresabschlussprüfung dar.<sup>2</sup> Für die Jahresabschlussprüfung bedeutet Wirtschaftlichkeit (im Folgenden auch als *Effizienz* bezeichnet), dass die wirtschaftlichsten Prüfungshandlungen einzusetzen und durchzuführen sind. Diese sind dadurch gekennzeichnet, dass die Differenz zwischen dem Urteilsgewinnungsbeitrag und den eingesetzten Mitteln im Vergleich zu anderen Prüfungshandlungen ein Maximum erzielt.<sup>3</sup> Für die Jahresabschlussprüfung bedeutet dies, dass die Abgabe eines hinreichend sicheren Prüfungsurteils zu minimalen Prüfungskosten zu erreichen ist.<sup>4</sup>

Die Notwendigkeit einer wirtschaftlichen Prüfungsdurchführung wird angesichts eines verschärften Wettbewerbs auf dem Prüfungsmarkt zusätzlich erhöht.<sup>5</sup> Darüber hinaus lässt auch die verstärkte Kapitalmarktorientierung der Unternehmen sowie die damit verbundene Notwendigkeit nach einer möglichst zeitnahen Finanzberichterstattung den Bedarf nach einer ökonomischen Ausgestaltung der Jahresabschlussprüfung steigen. Berücksichtigt man schließlich, dass der Umfang des Prüfungsstoffs nicht zuletzt aufgrund der rasanten Entwicklungen im IT-Bereich weiter zunimmt, so wird deutlich, dass die Entwicklung von kostengünstigeren Prüfungshandlungen einen wichtigen Wettbewerbsvorteil von Prüfungsgesellschaften darstellen kann.

Eine Möglichkeit zur Erhöhung der Effektivität und Effizienz von Jahresabschlussprüfungen liegt in der Verwendung von analytischen Prüfungshandlungen. Ihr Vorteil besteht darin, dass

---

<sup>1</sup> Vgl. etwa IDW PS 200.8-10, ISA 200.2.

<sup>2</sup> Vgl. IDW PS 200.9, ISA 300.2 f.

<sup>3</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 213.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 913-914.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 624. Ähnlich argumentiert auch QUICK (1999), S. 179 m.w.N.

sie keine einzelnen Geschäftsprozesse und Bestandsgrößen (sog. Einzelfallprüfungen), sondern agglomerierte Größen betrachten. Dabei wird unter Rückgriff auf bekannte oder erwartete Zusammenhänge in den jahresabschlussrelevanten Daten ein Ersatztatbestand entwickelt, der den zu prüfenden Werten gegenübergestellt wird und so Aufschluss über die Normenkonformität des Jahresabschlusses und des Lageberichts gibt. In der empirischen Prüfungsfor- schung wurde die Fehleraufdeckungsfähigkeit von analytischen Prüfungshandlungen bereits mehrfach nachgewiesen. HYLAS/ASHTON etwa stellten anhand einer Dokumentenanalyse fest, dass 37,4% aller wesentlichen Fehler anhand von analytischen Prüfungshandlungen enthüllt wurden.<sup>1</sup> Zu vergleichbaren Ergebnissen kamen BIGGS/MOCK. Sie zeigten, dass 41,5% aller wesentlichen Fehler anfänglich durch analytische Prüfungshandlungen angezeigt wurden.<sup>2</sup> WRIGHT/ASHTON erkannten, dass analytischen Prüfungshandlungen 15,5% der betragsmäßig hohen Fehler<sup>3</sup> identifizierten. Zudem stellten sie fest, dass etwa 50% der durch Einzelfallprü- fungen aufgedeckten Fehler auch durch analytische Prüfungshandlungen signalisiert worden wären.<sup>4</sup> Schließlich konnten auch FRASER/HATHERLY/LIN die Wirksamkeit von analytischen Prüfungshandlungen nachweisen. Sie stellten anhand einer Befragung fest, dass etwa 44,2% aller identifizierten Fehler durch analytische Prüfungshandlungen angezeigt wurden.<sup>5</sup> Darüber hinaus sind analytische Prüfungshandlungen häufig die einzige Möglichkeit, um bestimmte Fehler – mitunter auch Bilanzmanipulationen – im Jahresabschluss aufzudecken.<sup>6,7</sup>

Neben der Effektivität von analytischen Prüfungshandlungen ist auch ihre Effizienz von hoher Bedeutung für die Prüfungspraxis. Da analytische Prüfungshandlungen von der Durchführung zeit- und kostenintensiver Vergleiche zwischen Aufzeichnungen aus der Finanzbuchführung und den einzelnen zugrunde liegenden Belegen absehen und vielmehr die Konsistenz und

---

<sup>1</sup> Vgl. HYLAS/ASHTON (1982), S. 758.

<sup>2</sup> Vgl. BIGGS/WILD (1984), S. 75.

<sup>3</sup> Dabei handelt es sich um Fehler, die 20% des gesamten Materiality-Werts überstiegen.

<sup>4</sup> Vgl. WRIGHT/ASHTON (1989), S. 718.

<sup>5</sup> Vgl. FRASER/HATHERLY/LIN (1997), S. 43.

<sup>6</sup> Soll etwa die Vollständigkeit der Umsatzerlöse eines Handelsunternehmens beurteilt werden, so erlangt der Prüfer durch Einzelfallprüfung der erfassten Verkäufe keine Informationen darüber, ob das Risiko besteht, dass weitere Verkäufe vorlagen, jedoch nicht gebucht wurden. In derartigen Fällen bieten analytische Prü- fungshandlungen die einzige Möglichkeit, den Fehler durch eine Abweichung von einem entwickelten Er- wartungswert zu signalisieren. Eine analytische Prüfungshandlung, welche die Umsatzerlöse etwa ins Ver- hältnis zum Wareneinsatz des Handelsunternehmens setzt, könnte hier durchaus einen verlässlichen Prü- fungsnachweis erbringen. Vgl. SCOTT/WALLACE (1994), S. 64; KONRATH (2002), S. 124.

<sup>7</sup> Ein Beispiel für einen Bereich, in dem der Einsatz von analytischen Prüfungshandlungen zur Aufdeckung von Bilanzmanipulationen dem Einsatz von Detailprüfungen überlegen ist, bietet der sog. Salad Oil King- Case. Hier wurden die Speiseöltanks überwiegend mit Wasser gefüllt und nur zu einem geringen Teil mit Speiseöl. Da Öl aufgrund seiner Konsistenz an der Oberfläche schwimmt, konnte dieser Schwindel nicht durch augenscheinliche Inspektion aufgedeckt werden. Der Umstand, dass das manipulierende Unternehmen i.d.R. auch die entsprechenden Belege fingiert, erhöht die Wahrscheinlichkeit, dass Detailprüfungen nicht zur Aufdeckung der Manipulation geführt hätte. Analytische Prüfungshandlungen hingegen wären in diesem Fall

wirtschaftliche Plausibilität einer Gesamtheit von Geschäftsvorfällen und Bestandsgrößen untersuchen, kann der Prüfer schnell zu einer Einschätzung über die Normenkonformität des Jahresabschlusses und des Lageberichts gelangen. Analytische Prüfungshandlungen stellen somit eine kostengünstige Alternative zur Durchführung von Einzelfallprüfungen dar.

Ein Beispiel für die Vorteilhaftigkeit von analytischen Prüfungshandlungen liefert etwa die Prüfung der Umsatzerlöse eines Hotelunternehmens. Die Umsätze setzen sich in der Regel aus einer Vielzahl von Zimmerrechnungen mit relativ geringen Beträgen zusammen. Eine Einzelfallprüfung wäre in diesem Fall sehr mühsam und kostspielig. Allerdings ließe sich hier mittels analytischer Prüfungshandlungen eine durchaus präzise Schätzung für die Umsatzerlöse ermitteln, indem man den durchschnittlichen Zimmerpreis pro Zimmer auf Monatsbasis mit der monatlichen Belegungssrate und der Anzahl der Zimmer multiplizieren würde. Der Prüfer hätte auf diesem Wege eine relativ genaue Schätzung ohne großen Zeitaufwand erlangt, anhand derer er im Folgenden die ausgewiesenen Umsatzerlöse auf Plausibilität beurteilen könnte.<sup>1</sup>

Angesichts ihrer Effektivität und Effizienz können analytische Prüfungshandlungen einen bedeutenden Beitrag zur Erlangung von vertrauenswürdigen Prüfungsurteilen zu minimalen Kosten leisten. Demnach erscheint es nur konsequent, dass nationale und internationale Studien einen zunehmenden Einsatz von analytischen Prüfungshandlungen in der Prüfungspraxis verzeichnen.<sup>2</sup> Es ist also zu konstatieren, dass die Prüfungsgesellschaften den Nutzen analytischer Prüfungshandlungen erkannt haben und ihr Einsatz ein fester Bestandteil der Jahresabschlussprüfung geworden ist.

Umso mehr überrascht es jedoch, dass nur wenige (Struktur-)Vorgaben zur Durchführung von analytischen Prüfungshandlungen existieren. In der Prüfungsliteratur wird zwar eine Vielzahl von unterschiedlichen Verfahren analytischer Prüfungshandlungen diskutiert.<sup>3</sup> Diesen kommen jedoch ausschließlich eine initiiierende Bedeutung zu, da sie auf vorhandene wesentliche Abweichungen im Rahmen eines Soll-Ist-Vergleichs hinweisen (können). Die analytische

---

ein probates Mittel gewesen, um zu erkennen, dass der ausgewiesene Bestand an Speiseöl den weltweiten Bedarf nach Speiseöl übertraf. Vgl. WALLACE (1991), S. 256-257.

<sup>1</sup> Vgl. GILL/COSSERAT/LEUNG/CORAM (2001), S. 339.

<sup>2</sup> Vgl. für den nationalen Bereich GÄRTNER (1994), S. 169-198; für den internationalen Raum siehe etwa HOLDER (1983); BIGGS/WILD (1984); DAROCA/HOLDER (1985); CUSHING/LOEBBECKE (1986); TABOR/WILLIS (1985); AMEEN/STRAWSER (1994); JOHNSON/JOHNSON (1997); MAHATHEVAN (1997); FRASER/HATHERLY/LIN (1997); LIN/FRASER (2003); MULLIGAN/INGSTER (1999); SMITH/PSAROS/HOLMES (1999). Dabei wurde überwiegend festgestellt, dass die Prüfer vornehmlich einfache Verfahren, wie etwa Vorjahresvergleiche, Kennzahlen- und Trendanalysen verwenden und weniger die aufwendigen mathematisch-statistischen Verfahren zum Einsatz kommen (eine ausführliche Darstellung der Verfahren erfolgt unter Kap. 2.2).

<sup>3</sup> Vgl. hierzu ausführlich Kap. 2.2.

Prüfungshandlung ist damit jedoch nicht beendet. Vielmehr umfasst sie weiterhin die zur Identifikation des Abweichungsgrunds notwendigen prüferischen Aktivitäten. Die analytische Prüfungshandlung ist demnach erst abgeschlossen, wenn der Prüfer zu einem Urteil über den (vermuteten) Grund für die wesentliche Abweichung gelangt ist. Daher wird die Durchführung einer analytischen Prüfungshandlung in der Prüfungsliteratur auch als *Urteilsbildungsprozess* bezeichnet.<sup>1</sup>

Auch den auf den Einsatz von analytischen Prüfungshandlungen bezogenen nationalen und internationalen Prüfungsstandards<sup>2</sup> sind lediglich wenige Hinweise zur prüferischen Vorgehensweise zu entnehmen. So wird lediglich verlangt, dass der Prüfer bei einer festgestellten wesentlichen Abweichung zunächst die Unternehmensleitung zu befragen und die Antworten anschließend mit anderen Prüfungsnachweisen, die im Laufe der Prüfung erlangt werden, zu bestätigen hat. Sollte die Unternehmensleitung dabei keine Erklärungen geben (können) oder der Prüfer die Erklärungen als nicht ausreichend ansehen, so sind andere Prüfungshandlungen erforderlich.<sup>3</sup>

Vor dem Hintergrund der wenigen (Struktur-)Vorgaben ist fraglich, ob die Vorgehensweisen in der Prüfungspraxis das Potenzial von analytischen Prüfungshandlungen vollständig ausschöpfen. Diverse empirische Studien zur Durchführung analytischer Prüfungshandlungen lassen jedoch ein effektivitäts- und effizienzminderndes Verhalten der Prüfer vermuten. Die Ergebnisse weisen darauf hin, dass Prüfer auf Schwierigkeiten stoßen, wenn sie den Grund für eine festgestellte Abweichung identifizieren sollen. Bedenklich ist dabei insbesondere, dass es den Prüfern zum Teil nicht möglich war, die tatsächliche Fehlerursache für die festgestellte Abweichung zu generieren.<sup>4</sup> Zwar konnte dabei festgestellt werden, dass die Erfolgsquote bei erfahrenen Prüfern höher war als bei weniger erfahrenen Prüfern, dennoch war auch ihre Vorgehensweise durch unterschiedliche Mängel gekennzeichnet, was z.T. einer erfolgreichen Fehleraufdeckung entgegen stand. Darüber hinaus konnte in unterschiedlichen Studien nachgewiesen werden, dass die nach den Prüfungsstandards durchzuführende Befragung der Unternehmensleitung möglicherweise zu einer Beeinflussung der weiteren prüferischen Vorgehensweise führt und die Effektivität und Effizienz analytischer Prüfungshandlungen zusätzlich beeinträchtigen kann.<sup>5</sup>

---

<sup>1</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 171; MARTEN/QUICK/RUHNKE (2007), S. 297-298.

<sup>2</sup> IDW PS 312 bzw. ISA 520.

<sup>3</sup> Vgl. IDW PS 312.27, ISA 520.18.

<sup>4</sup> Vgl. etwa BEDARD/BIGGS (1991a), S. 636.

<sup>5</sup> Vgl. etwa BIERSTAKER/BEDARD/BIGGS (1999), S. 32-33.



Vor dem Hintergrund der wenigen – und sich möglicherweise negativ auswirkenden – Vorgaben ist es fraglich, ob analytische Prüfungshandlungen tatsächlich einen Beitrag zur Erlangung eines vertrauenswürdigen Prüfungsurteils zu minimalen Kosten leisten können. Sollte der Einsatz von analytischen Prüfungshandlungen aufgrund der mangelnden Fehleraufdeckungsfähigkeit vielmehr zur Abgabe inkorrektur Prüfungsurteile führen, so geht hiervon eine wesentliche Gefahr für die Effektivität – und mithin die Qualität – von Jahresabschlussprüfungen aus. Die Gewinnung von Vorgaben zur prüferischen Vorgehensweise scheint dringend notwendig.

Die Herleitung von zuverlässigen Vorgaben könnte sich vor dem Hintergrund der interpersonellen, vom jeweiligen Wissens- und Erfahrungsstand des Prüfers abhängigen Fehleraufdeckungsfähigkeit von analytischen Prüfungshandlungen weiterhin positiv auf die Effizienz von Jahresabschlussprüfungen auswirken. Weniger erfahrenen Prüfern könnte durch Vorgaben zu einer erfolgreichen und zügigen Durchführung von analytischen Prüfungshandlungen verholten werden. Da Prüfer mit wenigen Berufsjahren regelmäßig kostengünstiger sind als erfahrene Prüfer, könnten Vorgaben einen weiteren Beitrag zur Durchführung von wirtschaftlichen Jahresabschlussprüfungen leisten.

Damit die Effektivität und Effizienz von Jahresabschlussprüfungen weiterhin gewährleistet ist und möglicherweise gesteigert werden kann, hat sich die vorliegende Arbeit zum Ziel gesetzt, weit reichende Empfehlungen zur Durchführung von analytischen Prüfungshandlungen herauszuarbeiten. Hierbei werden insbesondere die bestehende Prüfungsliteratur sowie die umfassenden Erkenntnisse der empirischen Prüfungsforschung einer tiefer gehenden Analyse unterzogen. Um der Abhängigkeit des Fehleraufdeckungspotenzials analytischer Prüfungshandlungen von dem Erfahrungs- und Wissensstand der Prüfer (sog. interne Einflussfaktoren) ausreichend Rechnung zu tragen, erfolgt die Analyse aus einem verhaltenswissenschaftlichen Blickwinkel, wobei insbesondere auf die Erkenntnisse der Kognitionspsychologie zurückgegriffen wird. Weiterhin bietet eine verhaltenswissenschaftlich geprägte Analyse den Vorteil, dass weitere prüfungsprozessimmanente Faktoren – wie etwa die Befragung der Unternehmensleitung – identifiziert werden können, die sich positiv oder negativ auf die Effektivität und Effizienz von analytischen Prüfungshandlungen auswirken können (sog. externe Einflussfaktoren). Die Identifikation von externen Einflussfaktoren wird dabei zum Anlass genommen, um Empfehlungen zur Implementierung von Strukturvorgaben auszusprechen, die etwa durch Gesetze, Prüfungsstandards oder interne Verfahrensanweisungen und Prüfungshandbücher umgesetzt werden können. Die Feststellung von internen Einflussfaktoren wird hingegen

herangezogen, um Empfehlungen für die Personalweiter- und -ausbildung von Prüfungsgesellschaften auszusprechen.

Anhand der Darstellung des Untersuchungsgangs wird aufgezeigt, wie die zuvor formulierten Ziele erreicht werden sollen.

## 1.2 Gang der Untersuchung

Im Anschluss an den einleitenden Teil erfolgt zunächst eine grundsätzliche Unterteilung zwischen einem *deskriptiven Teil* und einem *empirischen Teil* der Arbeit. Der deskriptive Teil umfasst dabei die beiden folgenden Kap. 2 und 3. Den empirischen Teil der Arbeit bildet hingegen Kap. 4.

Kap. 2 beschäftigt sich mit der Anwendung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung. Hierbei ist zunächst der *Begriff* der analytischen Prüfungshandlung allgemein zu bestimmen (Abschnitt 2.1.1), um daraufhin einen ausführlichen Überblick über ihre *Anwendungsgebiete* im Rahmen des Prüfungsprozesses geben zu können (Abschnitt 2.1.2). Es wird gezeigt, dass analytische Prüfungshandlungen in der Phase der *Prüfungsplanung* (Abschnitt 2.1.2.2), der *Prüfungsdurchführung* (Abschnitt 2.1.2.3) und der *abschließenden Gesamtdurchsicht* (Abschnitt 2.1.2.4) zur Anwendung kommen können.

Anschließend erfolgt eine detaillierte Darstellung der Verfahren analytischer Prüfungshandlungen (Abschnitt 2.2). Die unterschiedlichen Verfahren werden nach der Art der Informationen klassifiziert, auf die zur Bildung eines Erwartungswerts zurückgegriffen wird. Dabei kann generell zwischen *Informationen aus der Vergangenheit* (Abschnitt 2.2.2), *Informationen aus dem aktuellen Geschäftsjahr* (Abschnitt 2.2.3) und *Informationen sowohl aus der Vergangenheit als auch aus dem aktuellen Geschäftsjahr* (Abschnitt 2.2.4) differenziert werden. Zur ersten Gruppe zählen der *einfache Vorjahresvergleich* (Abschnitt 2.2.2.1), die *einfachen Trendanalysen* (Abschnitt 2.2.2.2) sowie die *zeitbezogenen Kennzahlenanalysen* (Abschnitt 2.2.2.3). Zur zweiten Gruppe gehören die *Benchmark-Vergleiche* (Abschnitt 2.2.3.1) und die *einfachen Rechenmodelle* (Abschnitt 2.2.3.2). Die dritte Gruppe umfasst die *mathematisch-statistischen Verfahren* (Abschnitt 2.2.4.2). Die einzelnen Verfahren werden zunächst ausführlich dargestellt und anschließend kritisch gewürdigt, wobei insbesondere ihre Güte und ihre Eignung zur Erlangung von verlässlichen Prüfungsnachweisen diskutiert wird. Kap. 2 schließt mit der Skizzierung eines konventionellen *Ablaufmodells zur Durchführung von analytischen Prüfungshandlungen* (Abschnitt 2.3), wobei das prüferische Vorgehen in einzelne *Prüfungsschritte* zerlegt wird. Als zentrale Schritte werden dabei die *Bildung eines Erwar-*

*tungswerts* (Abschnitt 2.3.2.1), die *Festlegung von Toleranzgrenzen* (Abschnitt 2.3.2.2), der anschließende *Soll-Ist-Vergleich* (Abschnitt 2.3.2.3) sowie die *Beurteilung* des Ergebnisses des Vergleichs (Abschnitt 2.3.2.4) herausgearbeitet.

Da sich der Nutzen einer ausführlichen Darstellung und Würdigung der Verfahren analytischer Prüfungshandlungen (Abschnitt 2.2) jedoch auf die Ermittlung des Erwartungswerts und die Festlegung von Toleranzgrenzen beschränkt und auch die Ausführungen zu den weiteren Prüfungsschritten (Abschnitt 2.3) lediglich geringfügig zur Ableitung von Handlungsempfehlungen geeignet sind, richtet sich der Blick im anschließenden Kap. 3 auf die bisherigen *empirischen Forschungsaktivitäten* zur Durchführung von analytischen Prüfungshandlungen. Um jedoch die Vorgehensweise der empirischen Prüfungsforschung – und damit auch die Vorgehensweise im Rahmen der eigenen empirischen Studie in Kap. 4 – *wissenschaftstheoretisch* zu positionieren, werden zunächst drei allgemeine Abschnitte vorangestellt. Hierbei werden zuerst *Theoretische und Konzeptionelle Vorbemerkungen* (Abschnitt 3.1) getroffen, auf deren Basis *Weiterführende Überlegungen* (Abschnitt 3.2) getätigt werden, die schließlich zu der *Skizzierung eines erfahrungswissenschaftlichen Forschungsprogramms für das Prüfungswesen* (Abschnitt 3.2.1.2.2) führen. Es wird gezeigt, dass dieses Forschungsprogramm im Sinne eines „Wegweisers“ die einzelnen Schritte für empirische Forschungsaktivitäten vorgibt. Zentrale Schritte des Forschungsprogramms sind in der *Abbildung* des betrachteten *Problems* durch geeignete Modelle oder Theorien, der Entwicklung einer *Hypothese* (die eine Vermutung über das noch zu beobachtende prüferische Verhalten ausspricht) bzw. der Formulierung einer *Forschungsfrage*, der *Gewinnung und Auswertung von Daten* sowie der abschließenden *Bestätigung* oder *Verwerfung* der entwickelten Hypothesen bzw. der *Beantwortung* der formulierten Forschungsfragen zu sehen.

Um einen verhaltenswissenschaftlichen Blickwinkel einzunehmen, wird das Forschungsprogramm durch *Verhaltensorientierte Modelle zur Beschreibung des prüferischen Vorgehens* (Abschnitt 3.2.2) sowie den darauf ausgerichteten *Methoden zur Datengewinnung und -auswertung* (Abschnitt 3.4) validiert. Es stellt sich heraus, dass insbesondere das auf NEWELL/SIMON zurückgehende Informationsverarbeitungs- bzw. *Problemlösungsmodell* (Abschnitt 3.2.2.3) eine geeignete Grundlage zu Rekonstruktion des prüferischen Verhaltens bildet. Daher erfolgt eine ausführliche Darstellung der potenziellen *Problemlösungsstrategien* (Abschnitt 3.2.2.3.3) sowie der *Wissensstrukturen* (Abschnitt 3.2.2.3.4.2), mit denen sich Problemlösungsprozesse angemessen modellieren lassen. Anschließend wird ein *Ablaufmodell zur Durchführung von analytischen Prüfungshandlungen aus dem Blickwinkel des Problemlösungsansatzes* entwickelt (Abschnitt 3.3). Hierbei erfolgt eine Integration des in Ab-

schnitt 2.3 skizzierten konventionellen Ablaufmodells zur Durchführung analytischer Prüfungshandlungen und den in Abschnitt 3.2.2.3.3 vorgestellten Problemlösungsstrategien. Es zeigt sich, dass sich insbesondere die sog. *generate und test-Strategie* (Abschnitt 3.2.2.3.3.2.2.2) auf die Durchführung von analytischen Prüfungshandlungen übertragen lässt. Dabei ist das abgeleitete Ablaufmodell als ein *Fünf-Phasen-Modell* zu verstehen, bei dem der Prüfer zunächst eine auf den betrachteten Sachverhalt bezogene *Problemrepräsentation* aufbaut (erste Phase), um daraufhin *Hypothesen* für eine festgestellte Abweichung zu *generieren* (zweite Phase). Anschließend begibt sich der Prüfer auf die *Suche nach Informationen* (dritte Phase), vor deren Hintergrund er die zuvor generierten *Hypothesen bewertet* (vierte Phase), um abschließend eine *Prüfungsentscheidung* zu treffen (fünfte Phase).

Das in Abschnitt 3.3 deduzierte Ablaufmodell wird im Folgenden herangezogen, um die umfassenden *Erkenntnisse der empirischen Forschungsaktivitäten zur Durchführung von analytischen Prüfungshandlungen systematisch darzustellen* (Abschnitt 3.5). Dabei lassen sich die Studien der bisherigen Prüfungsforschung nach der Phase der *Problemrepräsentation* (Abschnitt 3.5.2), der *Hypothesengenerierung* (Abschnitt 3.5.3), der *Informationssuche* (Abschnitt 3.5.4), der *Hypothesenbewertung* (Abschnitt 3.5.5) und der abschließenden *Prüfungsentscheidung* (Abschnitt 3.5.6) unterteilen. Einer kurzen Skizzierung der jeweiligen Studien folgt die Darstellung ihrer Befunde. Anschließend werden die Erkenntnisse herangezogen, um zentrale Einflussfaktoren des prüferischen Vorgehens bei der Durchführung von analytischen Prüfungshandlungen zu identifizieren, auf deren Grundlage Handlungsempfehlungen zur Steigerung der Effektivität und Effizienz von analytischen Prüfungshandlungen abgeleitet werden.

Obwohl die o.a. empirischen Forschungsaktivitäten bereits eine geeignete Grundlage darstellen, um das in Abschnitt 3.3 deduzierte Ablaufmodell mit zahlreichen wichtigen Erkenntnissen zu validieren, existieren weitere offene Fragen zur Durchführung von analytischen Prüfungshandlungen. Insbesondere konnte keine abschließende Gewissheit darüber erlangt werden, ob die durch IDW PS 312.27 und ISA 520.18 empfohlene Befragung der Unternehmensleitung, falls eine wesentliche unerwartete Abweichung identifiziert wurde, zu einer nachhaltigen Beeinträchtigung der Effektivität und Effizienz von analytischen Prüfungshandlungen führt. Dieses nimmt der Autor zum Anlass, um eine eigene empirische Studie zur Durchführung von analytischen Prüfungshandlungen zu konzipieren. Dabei sollen entsprechend des in Abschnitt 3.2.1.2.2 abgeleiteten Forschungsprogramms zunächst eine Darstellung des zu untersuchenden Sachverhalts erfolgen (Abschnitt 4.1), um anschließend auf der Grundlage von vermuteten Ursache-Wirkungszusammenhängen *Hypothesen aufzustellen* und *Forschungs-*

*fragen zu formulieren* (Abschnitt 4.2). *Die Gewinnung* der notwendigen *Daten* erfolgt anhand eines *Experiments*, dessen *Aufbau* in Abschnitt 4.3 ausführlich dargestellt wird. Daraufhin werden die aufgestellten Hypothesen unter Rückgriff auf Methoden der *Datenauswertung* entweder bestätigt oder abgelehnt (Abschnitt 4.4). Anschließend werden die *Ergebnisse zusammengefasst* sowie die entsprechenden *Schlussfolgerungen* aufgezeigt (Abschnitt 4.5) und kritisch *diskutiert* (Abschnitt 4.6). Schließlich werden die *Grenzen* der empirischen Untersuchung skizziert (Abschnitt 4.7).

Das abschließende Kap. 5 der Arbeit ist einer *Zusammenfassung* sowie einem *Fazit* gewidmet.

## **2 Analytische Prüfungshandlungen als Gegenstand der Jahresabschlussprüfung**

### **2.1 Grundlegende Überlegungen**

#### **2.1.1 Begriffsbestimmung und Wesen analytischer Prüfungshandlungen**

Zur Erlangung eines hinreichend sicheren Gesamturteils über die Normenkonformität des Jahresabschlusses obliegt es dem Abschlussprüfer, Prüfungsnachweise (audit evidence) einzuholen, die ihn bei der Beurteilung der Normenkonformität einzelner Prüfungsgegenstände und/oder Prüffelder unterstützen.<sup>1</sup> Auf Grundlage der erlangten Prüfungsnachweise bildet der Prüfer diverse Prüfungsfeststellungen, die durch Aggregation zu einzelnen Prüfungsurteilen in Bezug auf das jeweilige Prüffeld führen.<sup>2</sup> Aus diesen (Teil-)Prüfungsurteilen leitet der Prüfer sein (Gesamt-)Prüfungsurteil über den Jahresabschluss in Ausprägung eines Bestätigungs- bzw. Versagungsvermerks ab.<sup>3</sup>

Zur Einholung von Prüfungsnachweisen stehen dem Prüfer unterschiedliche Prüfungshandlungen (audit procedures) zur Verfügung, wobei grundsätzlich zwischen der *Systemprüfung* einerseits und den *aussagebezogenen Prüfungshandlungen* andererseits differenziert werden kann. Während die Systemprüfung die Untersuchung der Zuverlässigkeit und Funktionsfähigkeit des internen Kontrollsystems des geprüften Unternehmens zum Gegenstand hat,<sup>4</sup> sind aussagebezogene Prüfungshandlungen darauf gerichtet, sich eine hinreichende Prüfungssicherheit darüber zu verschaffen, ob die in dem jeweiligen Jahresabschluss enthaltenen Aussagen des Unternehmens (sog. Abschlussausagen) den Anforderungen der geltenden Rechnungslegungsnormen genügen.<sup>5</sup>

Aussagebezogene Prüfungshandlungen lassen sich in *Einzelfallprüfungen* und *analytische Prüfungshandlungen* unterteilen.<sup>6</sup> Einzelfallprüfungen untersuchen unmittelbar die relevanten Einzelsachverhalte der Buchführung und des Jahresabschlusses, indem einzelne Geschäftsvorfälle und Bestandselemente geprüft werden. Daher werden diese u.a. auch als Detailprüfungen oder direkte Prüfungen bezeichnet.<sup>1</sup> Sie erlauben es dem Prüfer einen unmittelbaren Rückschluss auf die Normenkonformität einer vom Unterneh-

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 624.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 496-497.

<sup>3</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 658.

<sup>4</sup> Vgl. IDW PS 260.50.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 618.

<sup>6</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 83. Als Synonym wird hier auch der Begriff „Ergebnisprüfung“ aufgeführt.

men getätigten Abschlusssatzung zu ziehen.<sup>2, 3</sup> Analytische Prüfungshandlungen beschäftigen sich hingegen mit der Beurteilung von aggregierten Größen der Rechnungslegung. Aufgrund der Agglomeration eignen sich analytische Prüfungshandlungen nicht dazu, Aussagen über *einzelne* Geschäftsvorfälle und Bestände zu treffen. Sie dienen grundsätzlich der pauschalen Prüfung von *aggregierten* Größen, indem die Konsistenz und die ökonomische Plausibilität von Gesamtheiten von Geschäftsvorfällen und Bestandselementen beurteilt werden.<sup>4</sup>

Aufgrund der zuvor beschriebenen Eigenart von analytischen Prüfungshandlungen führt ihr Einsatz im Rahmen der Jahresabschlussprüfung zur Erlangung von *indirekten* Prüfungsnachweisen, welche im Allgemeinen in einem geringeren Maße zur Prüfungssicherheit beitragen als *direkte* Prüfungsnachweise.<sup>5, 6</sup> Bei indirekten Prüfungsnachweisen werden *Ersatztatbestände* herangezogen, um die Normenkonformität der im Jahresabschluss getroffenen Abschlusssatzungen beurteilen zu können.<sup>7</sup> Im Kontext von analytischen Prüfungshandlungen stellt dieser Ersatztatbestand ein vom Prüfer entwickelter Erwartungswert für eine bestimmte Jahresabschlussgröße dar, welcher anschließend dem tatsächlich im Jahresabschluss ausgewiesenen Buchwert gegenübergestellt und auf Abweichungen oder Übereinstimmungen hin untersucht wird. Bei der Bildung des Erwartungswerts greift der Prüfer im Wesentlichen auf seine speziellen Kenntnisse über den Mandanten, die Branche, in der dieser tätig ist, und die gesamtwirtschaftliche Entwicklung zurück.<sup>8</sup> Darüber hinaus ist ebenfalls notwendig, dass der Prüfer über ein Mindestmaß an rechnungslegungsspezifischem Wissen, etwa der doppelten Buchführung, verfügt, um die sachlogischen Zusammenhänge zwischen den einzelnen Jahresabschlusspositionen untereinander oder in Bezug zu einzelnen betrieblichen Daten erkennen und modellieren zu können.

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 228.

<sup>2</sup> Vgl. QUICK (2002), Sp. 1686.

<sup>3</sup> Beispielhaft kann hier die Inaugenscheinnahme der Vermögensgegenstände, etwa im Rahmen der Inventurbeobachtung, aufgeführt werden, welche dem Prüfer die Beurteilung der Existenz der im Jahresabschluss ausgewiesenen Vermögensgegenstände erlaubt. Ebenso führt der unmittelbare Vergleich zwischen Dokumenten (etwa Ein- und Ausgangsrechnungen) und den jeweiligen Aufzeichnungen in der Finanzbuchführung zu direkten Nachweisen in Bezug auf die Vollständigkeit und die Existenz bzw. den Eintritt von etwa im Rahmen der Verbindlichkeiten und Forderungen ausgewiesenen Geschäftsvorfällen.

<sup>4</sup> Vgl. QUICK (1999), S. 210.

<sup>5</sup> Vgl. KONRATH (2002), S. 113.

<sup>6</sup> Direkte Prüfungsnachweise werden durch den Einsatz von Detailprüfungen bzw. direkten Prüfungen erlangt.

<sup>7</sup> Vgl. QUICK (2002), Sp. 1685 sowie KAYADELEN (2006), S. 25.

<sup>8</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 297-298 sowie KAYADELEN (2006), S. 27.

Weitaus schwieriger als die Beschreibung der Eigenart von analytischen Prüfungshandlungen stellt sich die Aufstellung einer *eindeutigen Definition* dar. Dieses liegt wohl nicht zuletzt daran, dass es für diese Art des prüferischen Vorgehens eine Vielzahl von unterschiedlichen Begriffen gibt. Im deutschsprachigen Raum lassen sich häufig die Begriffe „Plausibilitätsbeurteilung“<sup>1</sup> oder „Plausibilitätsprüfung“<sup>2</sup> finden. Weitere Begriffe stellen die „Verprobung“<sup>3</sup>, die „Globalprüfung“<sup>4</sup> sowie die „Abstimmungsprüfung“ dar. Verwirrend ist allerdings auch, dass in der neueren Literatur häufig die Bezeichnung „analytische Prüfung“<sup>5</sup> zu finden ist. Vergleichbar verhält es sich in der angloamerikanischen Literatur. Hier stößt man einerseits auf die Begriffe „analytical review“<sup>6</sup> oder „analytical auditing“<sup>7</sup> und andererseits auf die Bezeichnungen „analytical review procedures“<sup>8</sup>, „analytical auditing procedures“<sup>9</sup> sowie „analytical evidence procedures“<sup>10</sup> oder einfach „analytical procedures“<sup>11</sup>. Um für den weiteren Verlauf der Untersuchung eine einheitliche terminologische Grundlage zu schaffen, wird im Folgenden aufgezeigt, welche der zuvor aufgeführten Begriffe synonym zu verwenden sind, und welche voneinander zu differenzieren sind.

Als Ausgangspunkt wird hierbei der vom IDW herausgegebene Prüfungsstandard 312 (IDW PS 312) herangezogen, der bereits in seinem Titel von analytischen Prüfungshandlungen spricht. In seiner Textziffer 5 heißt es anschließend:

*„Analytische Prüfungshandlungen sind Plausibilitätsbeurteilungen ...“*<sup>12</sup>

Fraglich ist allerdings, ob diese Formulierung eine inhaltliche Identität der beiden Begriffe zum Ausdruck bringen soll, oder ob „Plausibilitätsbeurteilung“ als Oberbegriff zu verstehen ist und demzufolge analytische Prüfungshandlungen eine Teilmenge von Plausibilitätsbeurteilungen darstellen. Letzterer Auffassung folgen GÄRTNER und MÜLLER. Beide gehen davon aus, dass der Begriff Plausibilitätsbeurteilung die eigentliche Prüfungsmethode, also die Art und Weise des prüferischen Vorgehens darstellt und mit dem Begriff analytische Prüfungshandlungen die verschiedenen Tätigkeiten des

---

<sup>1</sup> Vgl. QUICK (2002), Sp. 1685.

<sup>2</sup> Vgl. KROMMES (2006), S. 211.

<sup>3</sup> Vgl. QUICK (2002), Sp. 1685. Auch BUCHNER (1997), S. 234 verwendet den Begriff der Verprobung.

<sup>4</sup> Vgl. BUCHNER (1997), S. 232.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 293-294; KAYADELEN (2006), S. 25.

<sup>6</sup> Vgl. BLOCHER/WILLINGHAM (1985).

<sup>7</sup> Vgl. MCKEE (1989).

<sup>8</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 8-9.

<sup>9</sup> Vgl. GAUNTT/GLEZEN (1997).

<sup>10</sup> Vgl. KONRATH (2002), S. 119.

<sup>11</sup> Vgl. BLOCHER/PATTERSON (1996).

<sup>12</sup> IDW PS 312.5.



Prüfers im Rahmen der Ausführung dieser konkreten Prüfungsmethode zum Ausdruck kommen sollen, mit deren Hilfe er zu den einzelnen Prüfungsurteilen gelangt.<sup>1</sup> Diese Begriffsabgrenzung ist jedoch kritisch zu betrachten, da ihr bei genauerer Betrachtung lediglich ein sehr oberflächlicher Zusammenhang zu bescheinigen ist. Schaut man sich die eigentliche Bedeutung des Begriffs „analytisch“ bzw. „Analyse“ (griech. analysis = auflösen, zerlegen) an, so stellt man fest, dass hierbei zum einen ein komplexer Sachverhalt in seine einzelnen Elemente zerlegt wird und zum anderen die Beziehungen der Elemente zueinander festgestellt werden. Die verschiedenen, vermeintlich analytischen Tätigkeiten des Prüfers bei der Ausführung von Plausibilitätsbeurteilungen bestehen bei verkürzter Betrachtung hingegen darin, (1) einen Erwartungs- bzw. Sollwert zu generieren, (2) diesen Wert einem aus den Jahresabschlussdaten zu entnehmenden Ist-Wert gegenüberzustellen und (3) eine Abweichung bzw. Übereinstimmung zwischen diesen Werten festzustellen. Führt der Soll-Ist-Vergleich zur Feststellung einer Abweichung, so ist (4) deren Ursache zu eruieren. Folgt man der Ansicht der beiden oben zuvor genannten Autoren, so müsste man allen drei bzw. vier Schritten einen analytischen Charakter bzw. die Zerlegung eines Ganzen in seine Bestandteile attestieren. Dieses kann jedoch nur abgelehnt werden, da weder bei der Gegenüberstellung zweier Werte noch bei der Feststellung oder Nicht-Feststellung eines Unterschiedbetrags eine analytische Vorgehensweise zu erkennen ist. Lediglich dem ersten Schritt, der Ermittlung des Erwartungswerts, und dem möglichen vierten Schritt, der Feststellung der Abweichungsursache, sind eine analytische Vorgehensweise zuzusprechen, da in beiden Fällen regelmäßig die Erkennung von sachlogischen Zusammenhängen jahresabschlussrelevanter Da-

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 9; MÜLLER (1996), S. 7. Beide Autoren folgen damit der Auffassung von LÜCK (1998), S. 632 i.V.m. S. 635. Nach LÜCK handelt es sich bei Prüfungsmethoden um vom Prüfer angewandte Verfahren, mit deren Hilfe er die Normenkonformität feststellen und somit zu einem Prüfungsurteil gelangen will. Charakteristisch ist hierbei insbesondere die zielgerichtete Vorgehensweise (griech. methodos = Weg zu etwas, d.h. Art und Weise des Vorgehens, um zu einem Ziel zu gelangen). Prüfungshandlungen hingegen sind die verschiedenen prüferischen Tätigkeiten, die einen Soll-Ist-Vergleich einschließlich der Urteilsbildung erst ermöglichen. Insofern stellen Prüfungshandlungen eine Teilmenge von Prüfungsmethoden dar. In der Literatur werden die Begriffe „Prüfungsmethode“ und „Prüfungshandlung“ jedoch nicht immer einheitlich voneinander abgegrenzt. So vertreten etwa MARTEN/QUICK/RUHNKE (2006), S. 618 einerseits die Meinung, dass *Prüfungshandlungen* der Erlangung von Prüfungsnachweisen dienen und in Systemprüfungen und aussagebezogene Prüfungshandlungen bzw. Ergebnisprüfungen zu klassifizieren sind. Gleichzeitig sind die Autoren jedoch der Auffassung, dass mit der Risikoanalyse, der Systemprüfung und der Ergebnisprüfung drei grundsätzliche *Prüfungsmethoden* voneinander zu unterscheiden sind (vgl. MARTEN/QUICK/RUHNKE (2007), S. 293-294). Um derartigen Überschneidungen für die vorliegende Arbeit entgegenzuwirken, soll der Begriff „Prüfungsmethode“ bzw. „Methode“ in Anlehnung an LÜCK als Obergriff für sämtliche im Rahmen des Prüfungsprozesses eingesetzten Verfahren herangezogen werden. Diese werden in einer ersten Stufe in die Risikoanalyse und die Prüfungshandlungen zur Einholung von Prüfungsnachweisen unterteilt werden. Prüfungshandlungen werden in Anlehnung an MARTEN/QUICK/RUHNKE in einer zweiten Stufe in Systemprüfungen und aussagebezogene Prüfungshandlungen klassifiziert.

ten zueinander notwendig ist. Demzufolge ist den Auffassungen von GÄRTNER und MÜLLER nicht zu folgen.

Eine weitere Möglichkeit der Ober- und Unterbegriff-Interpretation besteht darin, dass neben analytischen Prüfungshandlungen weitere Arten von Plausibilitätsbeurteilungen existieren. Dieses ist jedoch grundsätzlich abzulehnen, da der gesamten einschlägigen Literatur zum Prüfungswesen nicht zu entnehmen ist, dass weitere Formen von Plausibilitätsbeurteilungen bestehen. Auch IDW PS 312 lässt in keiner Weise erkennen, dass eine derartige Abgrenzung dieser beiden Begriffe gewünscht ist.

Nachdem zuvor aufgezeigt wurde, dass die Auslegung des Begriffs „Plausibilitätsbeurteilung“ als Oberbegriff ausgeschlossen werden kann, erscheint es durchaus zweckmäßig, von dieser Interpretation Abstand zu nehmen und auf die eingangs genannte Interpretationsmöglichkeit zurückzugreifen. Demnach wäre davon auszugehen, dass die Formulierung des IDW eine inhaltliche Identität der beiden Begriffe konstatieren möchte und lediglich eine sprachliche Umformulierung vornimmt, um quasi einen höheren Verständlichkeitsgrad zu erreichen, ohne dabei jedoch einen materiellen Unterschied manifestieren zu wollen. Diese – die Gleichheit der beiden Begriffe unterstützende – Interpretation erscheint durchaus sinnvoll und besitzt daher auch für diese Untersuchung Geltung.<sup>1</sup>

Ebenfalls als Synonyme sind die beiden der steuerlichen Außenprüfung zuzuordnenden Begriffe „Verprobung“ und „Plausibilitätsprüfung“ zu verstehen.<sup>2</sup> Hierbei geht es analog zu den analytischen Prüfungshandlungen darum, dem vom geprüften Unternehmen ausgewiesenen Ist-Wert einen durch Kenntnis der funktionalen Beziehungen im Unternehmen oder Analogien entwickelten Erwartungswert gegenüberzustellen und auf seine Plausibilität zu untersuchen.<sup>3</sup>

Anders verhält es sich hingegen bei dem Begriff „Abstimmungsprüfung“, die lediglich auf die rein rechnerische Prüfung der korrekten Verarbeitung der Daten aus der Finanzbuchführung gerichtet ist und keine Untersuchung der wirtschaftlichen Plausibilität von Jahresabschlussdaten bezweckt.<sup>4</sup> Folglich ist dieser Begriff nicht mit dem der analyti-

---

<sup>1</sup> Gleicher Ansicht ist QUICK (2002), Sp. 1685, der konstatiert, dass Plausibilitätsbeurteilungen auch als analytische Prüfungshandlungen bezeichnet werden.

<sup>2</sup> Vgl. MÜLLER (1996), S. 7. m.w.N. sowie BUCHNER (1997), S. 234 in Bezug auf die Verprobung.

<sup>3</sup> Vgl. GÄRTNER (1994) S. 11 m.w.N.

<sup>4</sup> Vgl. BUCHNER (1997), S. 232.

schen Prüfungshandlung bzw. Plausibilitätsbeurteilung gleichzusetzen, da er erkennbar zu kurz greift.

Wesentlich schwieriger gestaltet sich dagegen die Abgrenzung von analytischen Prüfungen zu analytischen Prüfungshandlungen. In weiten Teilen des deutschsprachigen Schrifttums werden diese Begriffe offensichtlich synonym benutzt.<sup>1</sup> Darüber hinaus ist dem Verfasser keine deutschsprachige Quelle bekannt, die explizit um eine Differenzierung dieser beiden Begrifflichkeiten bemüht ist. Abhilfe verschafft derweil ein Blick in die angloamerikanische Literatur. Dort ordnet man den Begriffen *analytical review*<sup>2</sup> oder *analytical auditing*<sup>3</sup>, der wohl dem Begriff der analytischen Prüfungen am nächsten kommt, eine weitaus größere Bedeutung zu, als den in dieser Untersuchung zuvor beschrieben analytischen Prüfungshandlungen bzw. Plausibilitätsbeurteilungen. So subsumieren etwa BLOCHER/WILLINGHAM unter *analytical review* zwei grundsätzlich voneinander zu unterscheidende Methoden. Dabei handelt es sich einerseits um *risk analysis* und andererseits um *analytical review procedures*.<sup>4</sup> Erstgenannte haben zum Ziel, die Wahrscheinlichkeit wesentlicher Fehler im Jahresabschluss als Ganzes zu beurteilen. Zentrale Elemente nehmen hierbei die Beurteilung der Integrität des Managements oder der Going-Concern-Prämisse sowie das Suchen nach sog. *red flags*, wie etwa dem Weggang von Führungskräften, der Einführung eines erfolgsabhängigen Bonussystems für das Management, einem geplanten Börsengang oder aber auch der Änderung der Zahlungsvereinbarungen mit den Lieferanten.<sup>5</sup> Diese Maßnahmen sind eher der Einschätzung des mit dem Prüfungsauftrag verbundenen inhärenten Risikos<sup>6</sup> verbunden und regelmäßig zu Beginn einer Jahresabschlussprüfung im Rahmen der Prüfungsplanung durchzuführen. Da hier jedoch kein Erwartungswert zum Zwecke eines Soll-Ist-Vergleichs gebildet wird, ist festzuhalten, dass diese Form des *analytical reviews* keine analytische Prüfungshandlung im Sinne der vorliegenden Untersuchung darstellt. Viel-

---

<sup>1</sup> GÄRTNER (1994), MÜLLER (1996), QUICK (2002), WYSOCKI (2003) sowie IDW (2006) verwenden den Begriff der analytischen Prüfungshandlung, MARTEN/QUICK/RUHNKE (2003) sowie KAYADELEN (2006) verwenden hingegen den Begriff der analytischen Prüfung.

<sup>2</sup> Vgl. SMITH (1989), BLOCHER/WILLINGHAM (1985).

<sup>3</sup> Vgl. MCKEE (1989).

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 8-10. Für Letztere können die Begriffe „analytical evidence procedures“ oder einfach „analytical procedures“ als synonym erachtet werden.

<sup>5</sup> Einen ausführlichen Überblick über derartige Methoden geben BLOCHER/WILLINGHAM (1985), S. 17-46.

<sup>6</sup> Das inhärente Risiko wird definiert als die Wahrscheinlichkeit für das Auftreten von wesentlichen Fehlern unter der Annahme, dass es keine internen Kontrollen gibt. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 396.

mehr ist die Risikoanalyse konsequent von den sog. *Prüfungshandlungen* abzugrenzen, die ihrerseits auf die Erlangung von Prüfungsnachweisen gerichtet sind.<sup>1</sup>

Anders verhält es sich hingegen bei der zweiten Methode, den sog. *analytical review procedures*. Diese werden eingesetzt, um die Fehlerwahrscheinlichkeit in einzelnen Größen des zu prüfenden Jahresabschlusses beurteilen zu können, wobei auf quantitative Methoden, wie etwa die Kennzahlen- oder Trendanalyse, zurückgegriffen wird, um einen Erwartungswert zu entwickeln, der anschließend mit dem entsprechenden Buchwert des Unternehmens verglichen und auf Abweichungsursachen hin untersucht wird.<sup>2</sup> Es ist offenkundig, dass analytical review procedures in ihrer Zielsetzung und ihrem Inhalt den analytischen Prüfungshandlungen im Sinne der vorliegenden Untersuchung entsprechen.

Als Ergebnis kann folglich festgehalten werden, dass es sich bei dem Begriff „analytische Prüfungen“ analog zu „analytical review“ bzw. „analytical auditing“ um einen Oberbegriff für zwei unterschiedliche Prüfungsmethoden handelt, und zwar der Risikoanalyse und den analytischen Prüfungshandlungen. Für den weiteren Verlauf dieser Untersuchung werden ausschließlich Letztere als relevant erachtet.

## **2.1.2 Anwendungsgebiete und Ziele analytischer Prüfungshandlungen**

### **2.1.2.1 Vorbemerkung**

Sowohl nach IDW PS 312 als auch ISA 520 können analytische Prüfungshandlungen in verschiedenen Phasen einer Jahresabschlussprüfung zur Anwendung kommen, wobei ihre eigentliche Zielsetzung stark vom jeweiligen Zeitpunkt ihres Einsatzes innerhalb der Abschlussprüfung abhängig ist. In der Phase der *Prüfungsplanung* etwa werden sie in erster Linie zur Entwicklung einer Prüfungsstrategie eingesetzt.<sup>3</sup> Bei der eigentlichen *Prüfungsdurchführung* dienen sie der Erlangung von Prüfungsnachweisen, wohingegen ihr Einsatz zum Ende einer Prüfung den Prüfer im Rahmen der *abschließenden Ge-*

---

<sup>1</sup> Wie bereits zu Beginn des vorliegenden Abschnitts erläutert, können Prüfungshandlungen in *Systemprüfungen* einerseits und *aussagebezogene Prüfungshandlungen* bzw. *Ergebnisprüfungen* andererseits differenziert werden.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 9.

<sup>3</sup> Vgl. IDW PS 240.11.

*samtdurchsicht* unterstützt.<sup>1</sup> Im Folgenden sollen die Ziele analytischer Prüfungshandlungen mit Rücksicht auf die unterschiedlichen Phasen genauer untersucht werden.

### **2.1.2.2 Analytische Prüfungshandlungen im Rahmen der Prüfungsplanung**

#### **2.1.2.2.1 Ziele analytischer Prüfungshandlungen im Rahmen der Prüfungsplanung**

Wesentliches Anliegen der Phase der Prüfungsplanung ist die Erstellung eines Prüfungsprogramms. Hierzu ist es zunächst notwendig, eine Prüfungsstrategie zu entwickeln.<sup>2</sup> Der Einsatz analytischer Prüfungshandlungen zielt in diesem Zusammenhang darauf ab, den Prüfer bei

1. *der Vertiefung seiner Kenntnisse über die Geschäftstätigkeit des Mandanten,*
2. *der Identifizierung von kritischen Prüfungsgebieten und*
3. *der Planung der Art, des Umfangs und des zeitlichen Ablaufs von weiteren Prüfungshandlungen*

zu unterstützen.<sup>3</sup>

##### **2.1.2.2.1.1 Vertiefung der Kenntnisse über die Geschäftstätigkeit des Mandanten**

Nach IDW PS 312 und ISA 520 hat der Abschlussprüfer durch die Anwendung von analytischen Prüfungshandlungen in der Phase der Prüfungsplanung seine Kenntnisse über die Geschäftstätigkeit und das wirtschaftliche Umfeld des Unternehmens zu vertiefen.<sup>4</sup> Überaus fraglich ist allerdings, was diese auf den ersten Blick doch durchaus eindeutige Aussage tatsächlich zum Ausdruck bringen möchte. Denn wenn man sich einmal anschaut, was die überwiegend aus dem angloamerikanischen Raum stammende Literatur, teilweise jedoch auch die deutschsprachige Literatur, unter der Vertiefung der Kenntnisse über die Geschäftstätigkeit des Mandanten versteht, darf doch stark bezweifelt werden, ob dieses nicht der eigentlichen Definition von analytischen Prüfungshand-

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 25.

<sup>2</sup> Vgl. IDW PS 240.11.

<sup>3</sup> Vgl. IDW PS 312.17 und .19.

<sup>4</sup> Vgl. IDW PS 312.17.

lungen widerspricht. So werden analytische Prüfungshandlungen von zahlreichen Vertretern der Literatur als Instrument zur *Beurteilung der wirtschaftlichen Lage des Mandanten*, vergleichbar zur Vorgehensweise der *Jahresabschlussanalyse*, betrachtet.<sup>1</sup> Hiernach sollen sie den Prüfer bei der allgemeinen Risikobeurteilung unterstützen, da die wirtschaftliche Lage eines Unternehmens auf das Vorliegen eines erhöhten Fehlerrisikos hindeuten könne.<sup>2</sup> Zu denken wäre hier etwa an die Feststellung einer Verschlechterung der Liquidität oder Profitabilität anhand einschlägiger Kennzahlen.<sup>3</sup> Dem „Analysten“ könnte dies einen Hinweis darauf geben, dass das betrachtete Unternehmen Schwierigkeiten bei der Erfüllung seiner kurzfristigen Zahlungsverpflichtungen haben könnte.<sup>4</sup> Eine derartige Situation birgt die Gefahr, dass das Management dem Anreiz unterliegt, die schlechte Unternehmenslage zu verschleiern, indem bestimmte Jahresabschlusspositionen über- oder unterbewertet werden.<sup>5</sup>

Was auf den ersten Blick durchaus sinnvoll erscheint, ist bei genauerer Betrachtung jedoch nicht mehr mit dem eigentlichen Wesen analytischer Prüfungshandlungen zu vereinbaren. Hierfür spricht einerseits, dass die zuvor skizzierte Vorgehensweise bei der Durchführung von analytischen Prüfungshandlungen nur oberflächlich widergespiegelt wird. So werden zwar die ungeprüften Jahresabschlussdaten bzw. auf diesen basierende Kennzahlen mit den entsprechenden Werten des Vorjahrs bzw. der Vorjahre oder Branchenwerten verglichen und Rückschlüsse aus diesen Vergleichen oder Trends gezogen. Allerdings mangelt es vollends an der Ermittlung eines auf den Kenntnissen des Prüfers zurückzuführenden Erwartungswerts anhand dessen die ausgewiesenen Buchwerte des Mandanten auf ihre Plausibilität hin untersucht werden. Stattdessen erweckt eine Beurteilung der wirtschaftlichen Lage des Unternehmens unter Zuhilfenahme der ungeprüften Buchwerte den Eindruck, dass die zu prüfenden Jahresabschlussdaten als korrekt aufgefasst werden. Dieses eher der Jahresabschlussanalyse vergleichbare Vorgehen ist jedoch nicht als eine wie auch immer geartete Ausprägung analytischer Prüfungshandlungen aufzufassen, da die Aufgabe der Jahresabschlussprüfung eben darin besteht, die

---

<sup>1</sup> Für den angloamerikanischen Sprachraum siehe etwa MCKEE (1989), S. 23, 28, BLOCHER/WILLINGHAM (1985), S. 6, ARENS/ELDER/BEASLEY (2005), GRAY/MANSON (2000), S. 335. Für den deutschen Sprachraum siehe MÜLLER (1996), S. 25-26 und GÄRTNER (1994), S. 58-61. Bei KNECHEL (2001) fällt in diesem Zusammenhang auch der Begriff „Business Measurement“. Siehe KNECHEL (2001), S. 271.

<sup>2</sup> siehe GÄRTNER (1994), S. 59, MÜLLER (1996), S. 25.

<sup>3</sup> Eine ausführliche Übersicht für Liquiditäts- und Profitabilitätskennzahlen gibt etwa KNECHEL (2001), S. 277 oder ARENS/ELDER/BEASLEY (2005), S. 215-216.

<sup>4</sup> Vgl. KNECHEL (2001), S. 276 oder PORTER/SIMON/HATHERLY (2003), S. 230.

<sup>5</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 31.

Korrektheit der Daten mit hinreichender Sicherheit festzustellen.<sup>1</sup> Würde dieses bereits im Rahmen der Prüfungsplanung angenommen werden, so wäre der eigentliche Zweck der Prüfung fundamental in Frage gestellt. Es ist folglich hervorzuheben, dass analytische Prüfungshandlungen nicht der Beurteilung der wirtschaftlichen Lage des Mandanten dienen.

Was bedeutet nun aber, dass analytische Prüfungshandlungen in der Planungsphase die Kenntnisse des Prüfers über die Geschäftstätigkeit und wirtschaftlichen Verhältnisse des Mandanten vertiefen sollen? Um dieses beurteilen zu können, ist es angebracht, sich die eigentliche Bedeutung des Worts *Vertiefung* genauer anzuschauen. Denn um eine Vertiefung des Wissen, oder einfacher ausgedrückt ein detaillierteres Wissen über einen bestimmten Sachverhalt zu erreichen, muss bereits ein gewisses Fundament an Wissen vorhanden sein. Übertragen auf den hier betrachteten Fall kann es sich nur um das Wissen des Prüfers handeln, dass er sich im Vorfeld der Prüfung (etwa aus Gesprächen mit dem Mandanten), während des abgelaufenen Geschäftsjahrs (z.B. aus der Presse), aus bereits abgeschlossenen Prüfungen des gleichen Mandanten oder allgemein im Rahmen seiner bisherigen Prüfungstätigkeit angeeignet hat. Daher ist davon auszugehen, dass der Prüfer zum Zeitpunkt der Prüfungsplanung bereits über ein *breites* Wissen über den Mandanten und dessen Geschäfte verfügt. Darüber hinaus sollte der Prüfer auch über maßgebliche Transaktionen und Ereignisse im abgelaufenen Geschäftsjahr informiert sein, von denen anzunehmen ist, dass sie sich wesentlich auf die wirtschaftliche Lage des Unternehmens im zu prüfenden Geschäftsjahr ausgewirkt haben. Hierauf basierend sollte der Prüfer konkrete Erwartungen bezüglich der Jahresabschlusspositionen bilden können, die ihm als Grundlage für die Durchführung von analytischen Prüfungshandlungen dienen.<sup>2</sup> Sollte sich im Verlauf der Untersuchung jedoch herausstellen, dass sich die tatsächlichen Werte von den erwarteten Werten erheblich unterscheiden, so könnte der Prüfer davon ausgehen, dass sein Verständnis über den Mandanten und seine Geschäftstätigkeit auf unvollständigen oder fehlerhaften Informationen basiert und zu falschen Schlussfolgerungen hinsichtlich der zu erwartenden Jahresabschlusspositionen geführt hat.<sup>3</sup> Der Prüfer hat nun die Gelegenheit, auf bisher nicht bekannte Gegebenheiten der Geschäftstätigkeiten des Mandanten aufmerksam zu werden und seinen Wis-

---

<sup>1</sup> Vgl. HITZIG (2004), S. 32.

<sup>2</sup> So auch KROMMES (2006), S. 211.

<sup>3</sup> Zwar besteht darüber hinaus die Möglichkeit, dass der Grund für die Abweichung auch auf einen Fehler in den herangezogenen Daten des Mandanten zurück geführt werden könnte, womit der Prüfer die zugrunde liegende(n) Jahresabschlussposition(en) als kritische(s) Prüfungsgebiet(e) zu behandeln hät-

sensstand zu aktualisieren bzw. zu vertiefen.<sup>1</sup> Demnach helfen analytische Prüfungshandlungen in der Planungsphase dem Prüfer, die Adäquanz seines mandantenspezifischen Wissens zu beurteilen.<sup>2</sup>

#### **2.1.2.2.1.2 Identifikation von kritischen Prüfungsgebieten**

Neben der Vertiefung der Kenntnisse über den Mandanten besteht ein weiteres Ziel der Anwendung von analytischen Prüfungshandlungen in der Planungsphase vor allem auch darin, Fehlerrisiken zu beurteilen, um darauf basierend Prüfungsschwerpunkte festzustellen.<sup>3</sup> Prüfungsschwerpunkte sind Prüfungsgebiete, in denen die Wahrscheinlichkeit für das Vorliegen von wesentlichen Fehlern besonders hoch ist.<sup>4</sup> Dieses erhöhte Fehlerisiko führt dazu, dass diese Prüffelder mit einer höheren Intensität geprüft werden.<sup>5</sup> Im angloamerikanischen Raum spricht man in diesen Zusammenhang auch von „attention directing procedures“.<sup>6</sup> Hierbei geben unerwartete wesentliche Abweichungen von den gebildeten Erwartungswerten erste Hinweise auf potenzielle Fehlerrisiken im Jahresabschluss.<sup>7</sup>

Mit dem Erkennen kritischer Prüfungsgebiete sowie deren Festlegung als Prüfungsschwerpunkte versucht der Prüfer, eine besonders effektive und effiziente Jahresabschlussprüfung zu gewährleisten. Effektivität stellt ein Kriterium für das Erreichen einer angemessenen Prüfungsqualität dar, da sie die Fähigkeit zur Aufdeckung von wesentlichen Fehlern widerspiegelt. Effizienz orientiert sich hingegen an der Wirtschaftlichkeit einer Jahresabschlussprüfung. Der Einsatz von analytischen Prüfungshandlungen, die es dem Prüfer ermöglichen, sich auf diejenigen Prüfungsgebiete zu konzentrieren, die mit einer hohen Wahrscheinlichkeit Fehler enthalten, kann als besonders effektiv bezeichnet werden. Zudem versetzen analytische Prüfungshandlungen den Prüfer in die Lage, seine personellen und zeitlichen Kapazitäten besonders effizient zu verteilen, da der weitere Prüfungsprozess entsprechend der erkannten Fehlerrisiken geplant werden kann.<sup>1</sup> Insofern ist zu erkennen, dass die Identifikation von kritischen Prüfungsgebieten eng mit der

---

te. Dieses soll an dieser Stelle jedoch nicht weiter thematisiert werden, da dieses den Gegenstand des anschließenden Unterkapitels darstellt.

<sup>1</sup> Vgl. auch IDW PS 312.19.

<sup>2</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 57; PORTER/SIMON/HATHERLY (2003), S. 230.

<sup>3</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 295-297.

<sup>4</sup> Vgl. GÄRTNER (1994), S. 63 Fn. 48.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 645.

<sup>6</sup> Vgl. stellvertretend WRIGHT/ASHTON (1989), S. 710.

<sup>7</sup> Vgl. BLOCHER/WILLINGHAM (1984), S. 6.



Planung der Art, des Umfangs und des zeitlichen Ablaufs von weiteren Prüfungshandlungen verknüpft ist, welche den Gegenstand des folgenden Abschnitts bildet.

#### **2.1.2.2.1.3 Planung der Art, des Umfangs und des zeitlichen Ablaufs von weiteren Prüfungshandlungen**

Diese Zielsetzung von analytischen Prüfungshandlungen richtet sich auf die im Rahmen der Prüfungsplanung vorzunehmende *Prüfungsprogrammplanung* und *Zeitplanung*.<sup>2</sup> Während die Planung der Art und des Umfangs weiterer Prüfungshandlungen den Gegenstand der Prüfungsprogrammplanung bildet, wird der zeitliche Ablauf des weiteren Prüfungsprozesses im Rahmen der Zeitplanung terminiert. Allerdings ist darauf hinzuweisen, dass dieser Trennung lediglich eine formale Bedeutung beizumessen ist, da diese Bereiche in der Prüfungspraxis kaum eindeutig voneinander abgegrenzt werden können.<sup>3</sup> Insofern soll im Folgenden nur die Prüfungsprogrammplanung, hier insbesondere die Planung des Prüfungsumfangs betrachtet werden, da lediglich dieser Bereich Anknüpfungspunkte für analytische Prüfungshandlungen bietet.

Wie in Kap. 2.1.2.2.1.2 dargestellt, können analytische Prüfungshandlungen einen Hinweis darauf geben, in welchen Gebieten des Jahresabschlusses die Prüfungshandlungen aufgrund eines erhöhten Fehlerrisikos auszuweiten sind. Allerdings stellt dieses nur die eine Richtung dar, wie analytische Prüfungshandlungen die Planung des Umfangs weiterer Prüfungshandlungen beeinflussen können. Eine weitere Möglichkeit besteht darin, dass analytische Prüfungshandlungen ein geringes Fehlerrisiko signalisieren, indem die vom Prüfer gebildeten Erwartungswerte nicht oder nur unwesentlich von den ungeprüften Buchwerten des Mandanten abweichen. In derartigen Fällen wäre eine Reduzierung des Umfangs weiterer Prüfungshandlungen durchaus vertretbar.<sup>4</sup> Allerdings wird der Einsatz von analytischen Prüfungshandlungen zur Reduzierung weiterer Prüfungshandlungen in der Literatur zum Teil sehr kritisch betrachtet. Die aufgeführten Argumente

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 64.

<sup>2</sup> Keine explizite Berücksichtigung findet hier die ebenfalls im Rahmen der Prüfungsplanung vorzunehmende Personalplanung. Es ist jedoch davon auszugehen, dass die Personalplanung implizit im Rahmen der Prüfungsprogrammplanung berücksichtigt wird, da die Art und damit die Schwierigkeit bzw. Komplexität der geplanten Prüfungshandlungen gleichzeitig eine entsprechende Eignung des eingesetzten Personals bedingt. Einfache Belegprüfungen etwa können durch weniger qualifizierte Mitarbeiter durchgeführt werden. Demgegenüber ist beispielsweise die Prüfung von geschätzten bzw. beizulegenden Werten von höher qualifizierten Mitarbeitern vorzunehmen.

<sup>3</sup> So auch GÄRTNER (1994), S. 56-57.

<sup>4</sup> So auch MÜLLER (1996), S. 26.

gegen eine derartige Verwendung basieren auf den unterschiedlichen Auswirkungen der beiden Fehlerarten auf Effektivität und Effizienz der Jahresabschlussprüfung.

Grundsätzlich kann zwischen einem alpha-Fehler (Fehler erster Art) und einem beta-Fehler (Fehler zweiter Art) differenziert werden.<sup>1</sup> Als alpha-Fehler wird das Risiko bezeichnet, dass ein normenkonformes Prüffeld abgelehnt wird. Im hier vorliegenden Kontext bedeutet dies, dass ein „unkritisches“ Prüfungsgebiet als kritisch eingestuft und der Umfang der Prüfungshandlungen ausgedehnt wird. Für die Abschlussprüfung würde dies zwar keine negative Auswirkung auf die Prüfungsqualität bedeuten, aber der erhöhte Prüfungsaufwand würde eine nicht unerhebliche Minderung der Prüfungseffizienz mit sich bringen.<sup>2</sup> Anders verhält es sich hingegen bei dem beta-Fehler, der eine wesentlich ernstere Gefahr für die Prüfungsqualität darstellt. Quasi als „Gegenpol“ zum alpha-Fehler spiegelt der beta-Fehler das Risiko wider, dass ein normenwidriges Prüffeld vom Prüfer akzeptiert wird. Auf den Einsatz von analytischen Prüfungshandlungen übertragen bedeutet dies eine Reduzierung des Prüfungsumfangs, falls der Soll-Ist-Vergleich keine wesentliche Abweichung erkennen lässt. Somit steigt die Wahrscheinlichkeit, dass ein im Prüffeld enthaltener Fehler unentdeckt bleibt und die Effektivität der Jahresabschlussprüfung ist erheblich gefährdet.

Fraglich ist allerdings, ob die Beweiskraft der Ergebnisse von analytischen Prüfungshandlungen ausreicht, um die geforderte Prüfungssicherheit trotz eines eingeschränkten Prüfungsumfangs zu erreichen. Darüber hinaus berufen sich die Kritiker darauf, dass eine Abschlussprüfung in erste Linie auf die Vermeidung des beta-Fehlers gerichtet sein sollte, da sich hieraus wesentlich höhere Gefahren für den Prüfer ergeben.<sup>3</sup> Um diesem Interesse gerecht zu werden, sollten analytische Prüfungshandlungen lediglich zur Erlangung einer negativen Gewissheit (negative assurance) eingesetzt werden, d.h. der Prüfer sollte lediglich von der Existenz wesentlicher Fehler im Prüffeld ausgehen, falls eine unerwartete Abweichung festgestellt wurde. Ungeeignet wären analytische Prüfungshandlungen demnach zur Verschaffung einer positiven Gewissheit (positive assurance), d.h. der Prüfer sollte eine festgestellte Übereinstimmung zwischen dem Soll- und Istwert nicht zum Anlass nehmen, um das Risiko wesentlicher Fehler im Prüffeld

---

<sup>1</sup> Für eine ausführliche Beschreibung der beiden Fehlerarten soll an dieser Stelle auf MARTEN/QUICK/RUHNKE (2006), S. 23 und 133-134 verwiesen werden.

<sup>2</sup> Vgl. HOLDER/COLLMER (1980), S. 32.

<sup>3</sup> Vgl. etwa BLOCHER/WILLINGHAM (1985), S. 12 oder HOLDER/COLLMER (1980), S.32.

als derart gering einzuschätzen, als dass die Reduzierung des Prüfungsumfangs gerechtfertigt wäre.<sup>1</sup>

#### **2.1.2.2.2 Prüfungsgegenstände analytischer Prüfungshandlungen im Rahmen der Prüfungsplanung**

Wie zuvor festgestellt, richtet sich die Prüfungsplanung auf die Entwicklung einer allgemeinen Prüfungsstrategie, die sich in der Erstellung eines Prüfungsprogramms konkretisiert. Die prüferischen Tätigkeiten beziehen sich hier folglich nicht auf das Treffen von konkreten Prüfungsfeststellungen<sup>2</sup> hinsichtlich der Korrektheit der Jahresabschlussdaten. Dieses stellt vielmehr das Ziel der anschließenden Phase der Prüfungsdurchführung dar. Daher beziehen sich analytische Prüfungshandlungen in dieser noch sehr frühen Phase der Jahresabschlussprüfung im Allgemeinen auf hochverdichtete Unternehmensdaten. Der Prüfer hält sich in diesem Stadium regelmäßig auf der Ebene von Jahresabschlusspositionen auf. Detailliertes Datenmaterial, etwa auf Kontenebene, steht dem Prüfer hier noch nicht zur Verfügung.<sup>3</sup>

Da zur Durchführung von analytischen Prüfungshandlungen im Rahmen der Prüfungsplanung auf hochaggregierte Daten zurückgegriffen wird, fokussiert der Prüfer hier weniger spezielle Konten des zu prüfenden Unternehmens, als vielmehr allgemeine Unternehmenswerte („broad entity measures“<sup>4</sup>). Als Gegenstände kommen hier vornehmlich generelle Eigenschaften wie die Liquidität<sup>5</sup>, die Profitabilität<sup>6</sup> sowie die Kapitalstruktur<sup>7</sup> und die Vermögensstruktur<sup>8</sup> des Unternehmens in Betracht.<sup>9</sup> Stellt der Prüfer hier eine ungewöhnliche Abweichung im zu prüfenden Geschäftsjahr (etwa zum Vorjahr, vom

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 7, 10; BLOCHER (1983), S. 23 sowie HITZIG (2004), S. 35.

<sup>2</sup> Prüfungsfeststellungen sind begründete Schlussfolgerungen, die auf der Grundlage von angemessenen und ausreichenden Prüfungsnachweisen gezogen werden und das Prüfungsurteil stützen. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 610.

<sup>3</sup> Vgl. PORTER/SIMON/HATHERLY (2003), S. 229.

<sup>4</sup> PORTER/SIMON/HATHERLY (2003), S. 280.

<sup>5</sup> Hierzu könnte der Prüfer beispielsweise auf die unterschiedlichen *Liquiditätsgrade* zurückgreifen. Vgl. auch die Ausführungen in Kap. 2.2.2.3.4.1.

<sup>6</sup> Als geeignete Kennzahlen sind hier die *Eigenkapital-* sowie *Gesamtkapitalrentabilität* zu erwähnen. Vgl. hierzu auch die Ausführungen in Kap. 2.2.2.3.4.1.

<sup>7</sup> In Betracht kommt etwa die *Verbindlichkeitenintensität*. Weitere Ausführungen zu dieser Kennzahl finden sich in Kap. 2.2.2.3.4.2.4.

<sup>8</sup> In diesem Zusammenhang kann etwa auf die *Vorratsintensität* und *Forderungsintensität* verwiesen werden. Eine ausführliche Darstellung dieser beiden Kennzahlen beinhalten Kap. 2.2.2.3.4.2.2 sowie 2.2.2.3.4.2.3.

<sup>9</sup> Vgl. PORTER/SIMON/HATHERLY (2003), S. 230.

Trend in den Vorjahren oder von einem anderen Vergleichwert<sup>1)</sup> fest, so könnte er dieses als ein Indiz für ein erhöhtes Fehlerrisiko in einem der berücksichtigten Prüffelder auffassen. In Betracht kommen hier in der Regel Jahresabschlusspositionen wie etwa Forderungen, Verbindlichkeiten, Vorräte, Umsatzerlöse und -kosten.

### **2.1.2.3 Analytische Prüfungshandlungen im Rahmen der Prüfungsdurchführung**

#### **2.1.2.3.1 Ziele analytischer Prüfungshandlungen im Rahmen der Prüfungsdurchführung und Entscheidungskriterien ihres Einsatzes**

Als aussagebezogene Prüfungshandlung liegt das generelle Ziel von analytischen Prüfungshandlungen im Rahmen der Prüfungsdurchführung in der Leistung eines Beitrags zur Erlangung einer hinreichenden Prüfungssicherheit über die Normenkonformität der im Jahresabschluss enthaltenen Abschlusssaussagen.<sup>2, 3</sup> Damit stehen dem Prüfer neben Einzelfallprüfungen auch analytische Prüfungshandlungen oder eine Kombination aus beiden Methoden zur Verfügung.<sup>4</sup> Die Entscheidung darüber, ob analytische Prüfungshandlungen allein oder in Kombination mit Einzelfallprüfungen zur Erlangung von angemessenen Prüfungsnachweisen zum Einsatz kommen, ist dabei der eigenverantwortlichen Beurteilung der erwarteten Effektivität und Effizienz der zur Verfügung stehenden Prüfungshandlungen durch den Prüfer überlassen.<sup>1</sup> Nach IDW PS 312.21 hat der Prüfer dabei folgende Aspekte in sein Entscheidungskalkül einzubeziehen:

- Verlässlichkeit der Ergebnisse;
- Art des Unternehmens und die Möglichkeit, Informationen wieder in ihre Bestandteile zu zerlegen;
- Verfügbarkeit der Informationen;
- Zuverlässigkeit der Informationen;
- Relevanz der verfügbaren Informationen;

---

<sup>1</sup> Vgl. hierzu Kap. 2.2.2.1 bis 2.2.2.3 sowie 2.2.3.1.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 618-619. Vgl. hierzu auch Kap. 2.1.1.

<sup>3</sup> Da sich die durch analytische Prüfungshandlungen beurteilten Abschlusssaussagen regelmäßig auf Prüffelder der Jahresabschlussprüfung beziehen, dient die Anwendung von analytischen Prüfungshandlungen zugleich der Beurteilung der Normenkonformität von Prüffeldern.

<sup>4</sup> Vgl. hierzu Kap. 2.1.2.3.2.

- Vergleichbarkeit der verfügbaren Informationen;
- Erkenntnisse aus der Prüfungsplanung, der Systemprüfung und den Vorjahresprüfungen.

Im Folgenden wird eine ausführliche Erörterung dieser Aspekte vorgenommen.

#### **2.1.2.3.1.1 Verlässlichkeit der Ergebnisse analytischer Prüfungshandlungen**

Die Verlässlichkeit von analytischen Prüfungshandlungen hängt wesentlich mit der Qualität des vom Prüfer zu entwickelnden Erwartungswerts zusammen. In diesem Kontext spielt die Prognostizierbarkeit von Ergebnissen eine dominierende Rolle, die wiederum maßgeblich von der Gültigkeit zugrunde liegender sachlogischer Verknüpfungen abhängt.<sup>2</sup> Der Prüfer sollte daher ein Gespür dafür entwickeln, welche Ursache-Wirkungs-Zusammenhänge im Datenmaterial vorliegen, da bei nur scheinbar verknüpften Größen das Risiko besteht, dass falsche Schlussfolgerungen gezogen werden und die entwickelten Erwartungswerte nicht verlässlich sind.<sup>3</sup>

Anhaltspunkte für einen sachlogischen Zusammenhang und darauf aufbauend für die Vorhersagbarkeit von Ergebnissen bieten folgende Überlegungen:

- Buchungstechnische Zusammenhänge

Aufgrund der doppelten Buchführung bestehen bereits diverse Zusammenhänge zwischen unterschiedlichen Konten bzw. Prüffeldern (wie etwa zwischen den Umsatzerlösen und den Forderungen).

- Die Beständigkeit des Unternehmensumfelds

In einem beständigen Umfeld sind Zusammenhänge besser vorhersehbar als in einem unbeständigen oder dynamischen Umfeld.<sup>4</sup>

- Art des zu prüfenden Kontos bzw. Prüffelds

– Zusammenhänge innerhalb der Konten der Gewinn- und Verlustrechnung sind in der Regel besser vorhersehbar als Zusammenhänge in der Bilanz, da die der

---

<sup>1</sup> Vgl. IDW PS 312.20.

<sup>2</sup> Vgl. ISA 520.12a.

<sup>3</sup> Vgl. GÄRTNER (1994), S. 74.

<sup>4</sup> Vgl. BLOCHER/PATTERSON (1996), S. 55.

Gewinn- und Verlustrechnung immanente Zeitraumbetrachtung weniger anfällig für zufällige Schwankungen ist als die Zeitpunktbetrachtung der Bilanz.<sup>1</sup> Darüber hinaus weisen Daten der Gewinn- und Verlustrechnung einen stärkeren Bezug zu betrieblichen Daten auf, so dass dem Prüfer eine höhere Anzahl potenziell relevanter Daten zur Abbildung einer sachlogischen Verknüpfung zur Verfügung stehen.<sup>2, 3</sup> Betriebsdaten weisen weiterhin den Vorteil auf, dass es sich nicht um Daten aus der Finanzbuchführung handelt. Somit ermöglichen sie die Berücksichtigung sachlogischer Verknüpfungen auf Basis von unabhängigeren Informationen.<sup>4, 5</sup>

- Zusammenhänge von Konten bzw. Prüffeldern, deren Gestaltung dem Ermessen des Managements unterliegen, sind in der Regel ebenfalls schlechter vorhersehbar.<sup>6</sup> Für gewöhnlich handelt es sich hierbei um sog. *Individualprüffelder*, die im Wesentlichen außergewöhnliche und mit erheblichen Beurteilungs- und Gestaltungsmöglichkeiten gekennzeichnete Sachverhalte beinhalten.<sup>7</sup> Hiervon abzugrenzen sind die sog. *Massenprüffelder*, die durch ein großes Buchungsvolumen und eine systematische Verarbeitungsweise geprägt sind und aufgrund ihrer Homogenität und Objektivität besser vorhersehbar sind.<sup>8, 9</sup> Zu denken wäre hier

---

<sup>1</sup> Vgl. BLOCHER/PATTERSON (1996), S. 55.

<sup>2</sup> Vgl. HOLDER/COLLMER (1980), S. 31. Einen ausführlichen Überblick über derartige Zusammenhänge bieten BECKER/PETERSON (1964), S. 414-415.

<sup>3</sup> Dieses ist insbesondere der Fall, wenn die Gewinn- und Verlustrechnung nach dem Umsatzkostenverfahren ermittelt wird, da das Umsatzkostenverfahren den Betriebserfolg ermittelt, indem es die Umsatzerlöse der Periode den Herstellungskosten der abgesetzten Produkte und Leistungen zuzüglich den nicht zu den Herstellungskosten zählenden Gemeinkosten gegenüberstellt. Dieses setzt eine funktionierende Kostenstellen- und Kostenträgerrechnung voraus, die eine Aufschlüsselung der Kosten nach den Stellen Herstellung, allgemeine Verwaltung, Vertrieb und sonstige betriebliche Stellen erlaubt. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 490.

<sup>4</sup> Vgl. GÄRTNER (1994), S. 85.

<sup>5</sup> Darüber hinaus bietet sich der verstärkte Einsatz von analytischen Prüfungshandlungen in der Gewinn- und Verlustrechnung auch aufgrund ihrer engen Verknüpfung mit der Bilanz an. Jeder Position der Gewinn- und Verlustrechnung steht eine korrespondierende Bilanzposition gegenüber. Da im Rahmen der Prüfung der Bestandskonten der Bilanz in der Regel Einzelfallprüfungen zum Einsatz kommen, erscheint es nicht nur aus Kostengesichtspunkten sinnvoll, für die Beurteilung der korrespondierenden Konten der Gewinn- und Verlustrechnung analytische Prüfungshandlungen anstelle von Einzelfallprüfungen einzusetzen. Vgl. KELL/BOYNTON (1992), S. 327.

<sup>6</sup> Vgl. BLOCHER/PATTERSON (1996), S. 55.

<sup>7</sup> In Betracht kommen hier insbesondere Prüffelder, bei denen im Rahmen der Bewertung auf Schätzungen und Prognosen zurückzugreifen ist. Beispielhaft kann auf die Bewertung von Rückstellungen verwiesen werden. Nach internationalen Rechnungslegungsvorschriften kommen als Individualprüffelder insbesondere auch das Sachanlagevermögen, die immaterielle Vermögenswerten oder die Finanzanlagen in Betracht, wenn zur Bestimmung ihres Wertansatzes auf Bewertungsmodelle, wie etwa das Discounted Cash Flow Verfahren, zurückgegriffen wurde.

<sup>8</sup> Vgl. MÜLLER (1996), S. 29.

<sup>9</sup> Vgl. ISA 520.12a.

etwa an die Umsatzerlöse und Forderungen sowie an Verbindlichkeiten oder Vorräte.

Da eine zu prüfende Jahresabschlussgröße in der Regel von mehreren Faktoren beeinflusst wird, nimmt die Qualität des Erwartungswerts mit der Anzahl der erkannten und berücksichtigten Einflussfaktoren zu. Würde etwa bei der Beurteilung der Personalaufwendungen lediglich die Anzahl der Mitarbeiter berücksichtigt werden, so wäre die Qualität des Erwartungswerts als geringwertig zu beurteilen, da weitere wesentliche Einflussfaktoren, wie etwa das Lohn- und Gehaltsniveau oder die durchschnittliche Arbeitszeit, nicht einbezogen wären.<sup>1</sup> Ähnlich verhält es sich bei der Schätzung der Umsatzerlöse eines Hotelunternehmens. Die Schätzung der jährlichen Umsatzerlöse durch Multiplikation der durchschnittlichen Belegungsrate mit dem durchschnittlichen Zimmerpreis und der Anzahl der Zimmer weist sicherlich eine geringere Qualität auf als bei zusätzlicher Differenzierung nach Drei-, Zwei- oder Einbettzimmern.<sup>2</sup>

Des Weiteren hängt die Verlässlichkeit einer analytischen Prüfungshandlung nicht unwesentlich von dem angewandten Verfahren zur Ermittlung des Erwartungswerts ab.<sup>3</sup> Dabei stehen dem Prüfer neben komplexen statistischen Modellen, wie etwa der Regressions- oder Box-Jenkins-Analyse, auch einfache Vergleiche zur Verfügung. Letztgenannte weisen einen hohen Grad an Subjektivität auf und sind naturgemäß weniger verlässlich als statistische Modelle, die aufgrund ihrer immanenten Objektivität durch eine höhere Präzision gekennzeichnet sind.<sup>4</sup> Darüber hinaus weisen statistische Modelle wie etwa die Regressionsanalyse den Vorteil auf, dass sie mit der Standardabweichung automatisch einen Parameter zur Verfügung stellen, der den Prüfer bei der Beurteilung der Verlässlichkeit des gebildeten Erwartungswerts unterstützt.<sup>5</sup>

Schließlich hängt die Verlässlichkeit der Ergebnisse von analytischen Prüfungshandlungen auch von der zu beurteilenden Abschlussaussage ab.<sup>6</sup> So ist beispielsweise denkbar, dass wesentliche Fehler bei bestimmten Aussagen schwieriger durch Einzelfallprüfungen identifiziert werden können als durch analytische Prüfungshandlungen. Insbesondere bei der Beurteilung der Vollständigkeit eines Prüfungsgegenstands zeigen

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 31.

<sup>2</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 60.

<sup>3</sup> Eine ausführliche Darstellung sowie kritische Würdigung der unterschiedlichen Verfahren erfolgt in Kap. 2.2.

<sup>4</sup> Vgl. WALLACE (1983b), S. 22; BLOCHER (1983), S. 28, 30.

<sup>5</sup> Vgl. KINNEY/MCDANIEL (1996), S. 53.

<sup>6</sup> Vgl. ISA 520.12b.

sich analytische Prüfungshandlungen häufig den Detailprüfungen überlegen.<sup>1</sup> Die Aufdeckung von nicht erfassten Geschäftsvorfällen durch den Einsatz von Einzelfallprüfungen erweist sich als äußerst problematisch, da dem Prüfer in diesen Fällen keine geeignete Grundlage zur Verfügung steht.<sup>2</sup>

#### **2.1.2.3.1.2 Art des Unternehmens und die Möglichkeit, Informationen wieder in ihre Bestandteile zu zerlegen**

Die Fehleraufdeckungsfähigkeit von analytischen Prüfungshandlungen hängt auch von der Detailtiefe des zur Entwicklung eines Erwartungswerts herangezogenen Datenmaterials ab.<sup>3</sup> Greift der Prüfer auf hoch verdichtete Daten zurück, so besteht die Gefahr, dass wesentliche Fehler nicht erkannt werden, wenn sich ihre gegenläufigen Auswirkungen neutralisieren.<sup>4</sup> Daher sollte der Prüfer möglichst detailliertes Datenmaterial verwenden, etwa bezogen auf die einzelnen Geschäftsbereiche bzw. Produktlinien und nicht auf das gesamte Unternehmen und/oder bezogen auf Monatsdaten und nicht auf das gesamte Geschäftsjahr.<sup>5</sup> Der Nutzen des Einsatzes von detailliertem Datenmaterial hängt sicherlich auch von der Art des zu prüfenden Unternehmens ab. Handelt es sich etwa um Unternehmen mit sehr unterschiedlichen Geschäftsbereichen oder hoher Produktdiversifikation, ist die Verwendung von weniger aggregierten Daten sinnvoller als bei sehr homogenen Unternehmen. Allerdings ist einschränkend festzuhalten, dass dem Prüfer nicht immer detaillierte Daten zur Verfügung gestellt werden (können).

#### **2.1.2.3.1.3 Verfügbarkeit der Informationen**

Für die Anwendung von analytischen Prüfungshandlungen ist es zwingend notwendig, dass dem Abschlussprüfer die notwendigen Daten auch zur Verfügung gestellt werden können. Der Prüfer sollte sich vergewissern, ob das betroffene Unternehmen entsprechende Finanzinformationen, wie etwa Budgets oder Planrechnungen, oder betriebliche Informationen, wie etwa Produktions- oder Verkaufszahlen, aufbereitet hat. In diesem Zusammenhang wird noch darauf hingewiesen, dass die zu prüfende Unternehmensor-

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 28; BLOCHER/PATTERSON (1996), S. 55.

<sup>2</sup> Vgl. KONRATH (2002), S. 123-124.

<sup>3</sup> Vgl. ISA 520.12e.

<sup>4</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 60; BLOCHER/PATTERSON (1996), S. 55.

<sup>5</sup> Vgl. SCOTT/WALLACE (1994), S. 64.



ganisation/-struktur nicht nur geeignet sein muss, um detaillierte Daten in die Entwicklung eines Sollwerts einfließen zu lassen, vielmehr muss das Unternehmen diese Informationen auch bereitstellen können und wollen.

#### **2.1.2.3.1.4 Zuverlässigkeit der Informationen**

Eine der zentralen Voraussetzungen für die Effektivität und Effizienz von analytischen Prüfungen ist die Zuverlässigkeit der für die Bildung des Erwartungswerts herangezogenen Informationen, da nur zuverlässige Daten die Grundlage für aussagekräftige Prüfungsnachweise bilden können.<sup>1</sup> Daher sollte sich der Prüfer vor der Durchführung analytischer Prüfungshandlungen zunächst von der Richtigkeit der heranzuziehenden Daten überzeugen. Aufgrund der Vielfältigkeit der zur Verfügung stehenden Daten hat der Prüfer eine Vielzahl von Aspekten zu berücksichtigen, um eine Abschätzung ihrer Zuverlässigkeit gewissenhaft vornehmen zu können. So kann er etwa auf Daten der Vorperioden, auf Budgets und Planzahlen, aktuelle Jahresabschluss- oder betriebliche Daten sowie Branchen- oder gesamtwirtschaftliche Daten zurückgreifen. Dabei können die Daten entweder aus internen oder externen Quellen stammen. Die Zuverlässigkeit von Daten aus internen Quellen hängt im Wesentlichen von der Leistungsfähigkeit des internen Kontrollsystems ab,<sup>2</sup> die Zuverlässigkeit von Daten aus externen Quellen hingegen von der Vertrauenswürdigkeit der Datenquelle.<sup>3</sup> Daten aus externen Quellen weisen aufgrund ihrer Unabhängigkeit zum Prüfungsgegenstand grundsätzlich eine höhere Zuverlässigkeit auf als Daten aus Quellen innerhalb des Unternehmens.<sup>4</sup> In der Regel wird der Prüfer seinen Erwartungswert jedoch nicht ausschließlich auf der Grundlage von Daten aus externen Quellen generieren können. Daher sollte er bei der Datenbeschaffung innerhalb des geprüften Unternehmens zumindest darauf achten, dass die Quellen unabhängig von den für den Prüfungsgegenstand verantwortlichen Personen sind.<sup>5</sup> In diesem Zusammenhang ist auch zu berücksichtigen, dass bereits geprüften Daten, etwa aus Vorjahren, eine höhere Zuverlässigkeit zugeordnet werden kann, als ungeprüften Daten.<sup>1</sup> Zudem sollte sich der Prüfer darüber bewusst sein, dass betriebliche Daten im Vergleich zu Finanzdaten zum Teil keinen oder geringeren internen Kontrollen unter-

---

<sup>1</sup> Vgl. ISA 520.12c.

<sup>2</sup> Da dieser Aspekt unter Kap. 2.1.2.3.1.7 noch ausführlicher diskutiert wird, erfolgt an dieser Stelle keine weitere Erörterung.

<sup>3</sup> Siehe hierzu BLOCHER/PATTERSON (1996), S. 55.

<sup>4</sup> Vgl. ISA 520.12c (a).

<sup>5</sup> Vgl. MÜLLER (1996), S. 30.

liegen, was ihre Zuverlässigkeit erheblich beeinträchtigt.<sup>2</sup> Schließlich ist die Zuverlässigkeit des in die Entwicklung des Erwartungswerts eingeflossenen Datenmaterials als höher einzuschätzen, wenn diese aus unterschiedlichen, voneinander unabhängigen externen oder internen Quellen stammen.<sup>3</sup>

#### **2.1.2.3.1.5 Relevanz der verfügbaren Informationen**

Eine weitere nicht zu vernachlässigende Voraussetzung für die Effektivität und Effizienz von analytischen Prüfungshandlungen ist die Relevanz der einfließenden Daten. Falls der Prüfer auf Informationen interner Quellen zurückgreift, sollte er sich auf Daten konzentrieren, die vom Management als entscheidende Einflussfaktoren für den Unternehmenserfolg erachtet und als Grundlage für unternehmerische Entscheidungen herangezogen werden.<sup>4</sup> Verwendet der Prüfer bei der Ermittlung eines Erwartungswerts etwa Budgets oder Planzahlen, so sollte er sich darüber vergewissern, ob es sich hierbei um realistische Zielgrößen handelt, oder ob das Unternehmen diese Vorgaben als Anreizinstrumente eingesetzt hat.<sup>5</sup> Letztere stellen kaum geeignete Daten zum Zwecke von analytischen Prüfungshandlungen dar.

#### **2.1.2.3.1.6 Vergleichbarkeit der verfügbaren Informationen**

Greift der Prüfer bei der Beurteilung der Plausibilität der zu prüfenden Jahresabschlussdaten auf Branchendaten zurück, so hängt die Wirksamkeit seines Vorgehens entscheidend von der Vergleichbarkeit der herangezogenen Daten mit den Daten des zu beurteilenden Unternehmens ab. In diesem Zusammenhang könnte es notwendig sein, dass etwa Branchendurchschnittswerte modifiziert werden, um sie mit den entsprechenden Werten des betrachteten Unternehmens vergleichen zu können, da dieses u.U. Spezialprodukte herstellt oder vertreibt.<sup>6</sup> Werden hingegen Werte von einzelnen Unternehmen

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 76 m.w.N.

<sup>2</sup> Vgl. WALLACE (1983a), S. 32; BLOCHER (1983), S. 24. Auch ISA 520.12c (d) weist darauf hin, dass sich der Prüfer bei der Verwendung von betrieblichen Daten zunächst darüber zu vergewissern hat, ob deren Entstehung ausreichenden Kontrollen unterliegt. ISA 520.12d gibt ein Beispiel wie dieses gleichzeitig mit der Prüfung des internen Kontrollsystems geschehen kann.

<sup>3</sup> Vgl. SCOTT/WALLACE (1994), S. 64.

<sup>4</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 60.

<sup>5</sup> Vgl. ISA 520.12c (c).

<sup>6</sup> Vgl. ISA 520.12c (b).

herangezogen, kommt es im Wesentlichen darauf an, dass dieses sog. Benchmarkunternehmen eine vergleichbare Unternehmensorganisation und -struktur aufweist.

#### **2.1.2.3.1.7     Erkenntnisse aus der Prüfungsplanung, den Vorjahresprüfungen und der Systemprüfung**

Im Rahmen der Entscheidung, ob sich der Prüfer verstärkt auf den Einsatz und die Ergebnisse von analytischen Prüfungshandlungen stützen kann, sollten insbesondere auch Erkenntnisse aus der Planungsphase, den Prüfungen der Vorjahre und der Beurteilung der Leistungsfähigkeit des internen Kontrollsystems einfließen.

Ist der Prüfer während der Prüfungsplanung etwa auf Prüffelder mit wesentlichen Fehlerrisiken gestoßen, sollte er in diesen Bereichen verstärkt Einzelfallprüfungen vornehmen, um hinreichend sichere und genaue Informationen über das tatsächliche Fehlerausmaß zu erhalten.<sup>1</sup> Aus vergleichbaren Überlegungen stellen auch die Arbeitspapiere der Vorjahresprüfungen eine wichtige Quelle für erhöhte Fehlerrisiken dar, da sie den Prüfer darüber informieren, in welchen Bereichen wesentliche Nachbuchungen vorgenommen wurden und welche Ursachen diesen zugrunde lagen.<sup>2</sup> Schließlich sollte sich der Einsatz analytischer Prüfungshandlungen auch nach den Ergebnissen der Systemprüfung richten.<sup>3</sup> Schwächen des internen Kontrollsystems führen zu einem erhöhten Kontrollrisiko und damit zu einer erhöhten Wahrscheinlichkeit für das Vorliegen von wesentlichen Fehlern im Jahresabschluss. In derartigen Fällen sollte sich der Prüfer weniger auf analytische Prüfungshandlungen als auf Einzelfallprüfungen verlassen, da diese eine höhere Beweiskraft besitzen.<sup>4</sup>

#### **2.1.2.3.2     Arten analytischer Prüfungshandlungen im Rahmen der Prüfungsdurchführung**

Analytische Prüfungshandlungen im Rahmen der Prüfungsdurchführung lassen sich in zwei Kategorien einteilen.<sup>5</sup> Bei der ersten Kategorie handelt es sich um den Einzelfallprüfungen „vorgeschaltete“ Test. Sie finden ihren Einsatz in Prüfungsgebieten, die un-

---

<sup>1</sup> Vgl. ISA 520.12b (a).

<sup>2</sup> Vgl. etwa SPRINKLE/TUBBS (1998), S. 475.

<sup>3</sup> Vgl. ISA 520.12d.

<sup>4</sup> Vgl. KINNEY/MCDANIEL (1996), S. 53.

<sup>5</sup> Vgl. im Folgenden PORTER/SIMON/HATHERLY (2003), S. 282.

abhängig von den Ergebnissen der analytischen Prüfungshandlung die Vornahme weiterer umfangreicher (Einzelfall-)Prüfungen erfordern (etwa Vorräte oder Forderungen). Hierbei unterstützen sie den Prüfer bei der Entscheidung, in welchem Umfang weitere Einzelfallprüfungen durchzuführen sind. Erscheint der ausgewiesene Buchwert plausibel und ist die Gefahr wesentlicher Fehler folglich als gering einzuschätzen, kann der Umfang weiterer (Einzelfall-)Prüfungen eingeschränkt werden.<sup>1</sup> Weisen Ergebnisse der analytischen Prüfungshandlung hingegen Inkonsistenzen im ausgewiesenen Buchwert aus und ist das Fehlerrisiko mithin erhöht, so ist der Umfang weiterer (Einzelfall-)Prüfungen zu erhöhen. Im angloamerikanischen Sprachraum wird diese Gruppe analytischer Prüfungshandlungen auch als „preliminary analytical procedures“<sup>2</sup> bezeichnet.

Die zweite Kategorie wird indes in solchen Prüfungsgebieten eingesetzt, die generell einen geringeren Prüfungsumfang erfordern („analytical procedures as complete test“).<sup>3</sup> Erscheinen die Ergebnisse plausibel, kann auf die Vornahme weiterer Prüfungshandlungen vollends verzichtet werden.<sup>4</sup> Sind hingegen Inkonsistenzen ersichtlich, sind auch hier weitere Prüfungshandlungen vorzunehmen.

Prüfungsgebiete, in denen der Einsatz analytischer Prüfungshandlungen vollständig ausreichen könnte, stellen etwa die Zinsaufwendungen oder das Konto „Gutschriften und Warenrückläufe“ dar. Beide Konten weisen einen nachvollziehbaren sachlogischen Zusammenhang zu einem anderen Konto auf (Umsatzerlöse bzw. Bankverbindlichkeiten). Hat sich in der Vergangenheit etwa ein stabiles Verhältnis zwischen den Umsatzerlösen und den Rückläufen eingestellt, so könnte der Prüfer auch für das aktuelle Geschäftsjahr die gleiche Relation unterstellen, falls ihm keine Informationen vorliegen, die gegen diese Annahme sprechen. Stellt der Prüfer mittels analytischer Prüfungshandlungen fest, dass das Verhältnis zwischen den Umsatzerlösen und den Rückläufen im aktuellen Geschäftsjahr nicht wesentlich vom Verhältnis des Vorjahrs abweicht, so könnte er dieses als ausreichenden Prüfungsnachweis für einen normenkonformen Buchwert ansehen.

Ein vergleichbarer Prüfungsnachweis wäre bei der Prüfung der Zinsaufwendungen denkbar. Hier könnte der erwartete Zinsaufwand durch das Produkt aus den durchschnittlichen Bankverbindlichkeiten und dem durchschnittlichen Zinssatz ermittelt wer-

---

<sup>1</sup> Dieses sieht auch IDW PS 312.11 vor. Vgl. hierzu auch IDW (2006), R 395.

<sup>2</sup> Vgl. etwa HARPER/STRAWSER/TANG (1990), S. 115; GAUNTT/GLEZEN (1996), S. 61; GLOVER/JIAMBALVO/KENNEDY (2000), S. 27.

<sup>3</sup> Vgl. im Folgenden PORTER/SIMON/HATHERLY (2003), S. 282.

<sup>4</sup> Auch eine derartige Vorgehensweise ist durch IDW PS 312.11 geregelt. Vgl. hierzu auch IDW (2006), R 395.

den. Weist der tatsächlich ausgewiesene Buchwert keine wesentliche Differenz zum erwarteten Wert auf, so könnte der Prüfer auch hier allein aufgrund der Vornahme analytischer Prüfungshandlungen davon ausgehen, dass das Prüfungsgebiet keine wesentlichen Fehler umfasst.

#### **2.1.2.3.3 Prüfungsgegenstände analytischer Prüfungshandlungen im Rahmen der Prüfungsdurchführung**

Im Gegensatz zur Prüfungsplanung, wo sich analytische Prüfungshandlungen im Wesentlichen auf hoch aggregierte Unternehmensdaten bzw. auf der Ebene von Jahresabschlusspositionen beschränken, richten sich analytische Prüfungshandlungen im Rahmen der Prüfungsdurchführung bereits auf einzelne Konten oder bestimmte Gruppen von Transaktionen und damit auf Daten mit einem geringeren Aggregationsgrad.<sup>1</sup> Da diese Prüfungshandlungen bereits den Charakter einer Tiefenuntersuchung („in-depth examination“<sup>2</sup>) aufweisen, werden sie mitunter als *specific analytical procedures* bezeichnet. Analytische Prüfungshandlungen, die im Rahmen der Prüfungsplanung eingesetzt werden, nennt man demgegenüber auch *overall analytical procedures*.<sup>3</sup>

Zur Durchführung von *specific analytical procedures* eignet es sich insbesondere, die Unternehmensdaten in sachlicher, zeitlicher oder geographischer Hinsicht aufzubereiten. So wäre im Rahmen der Prüfung der Vorratsbestände etwa denkbar, dass das Verhältnis der Lagerbestände zu den Umsatzkosten nach Produkten und/oder Lagerstandorten berechnet wird.<sup>4</sup> Zur Prüfung der Umsatzerlöse besteht hingegen die Möglichkeit, die Umsätze pro Produktgruppe und/oder Monat zu kalkulieren. Werden dem Prüfer entsprechende Daten zur Verfügung gestellt, so könnte er die Umsatzerlöse pro Kunden, pro Verkaufsmitarbeiter, pro Verkaufstransaktion oder pro Quadratmeterzahl der entsprechenden Vertriebsstätten berücksichtigen. Schließlich wäre es auch möglich, dass zur Prüfung der Umsatzkosten die anteiligen Material- und Fertigungseinzel- sowie

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 7.

<sup>2</sup> BLOCHER/WILLINGHAM (1985), S. 7.

<sup>3</sup> Vgl. PORTER/SIMON/HATHERLY (2003), S. 280.

<sup>4</sup> Eine ausführliche Darstellung der Kennzahlenanalyse als ein potenzielles Verfahren analytischer Prüfungshandlungen erfolgt in Kap. 2.2.2.3.

-gemeinkosten pro Produkt und/oder Monat ins Verhältnis zu den gesamten Umsatzkosten pro Produkt und/oder Monat gesetzt werden.<sup>1</sup>

#### **2.1.2.4 Analytische Prüfungshandlungen im Rahmen der abschließenden Gesamtdurchsicht**

##### **2.1.2.4.1 Ziele analytischer Prüfungshandlungen in der abschließenden Gesamtdurchsicht**

###### **2.1.2.4.1.1 Vorbemerkungen**

Analytische Prüfungshandlungen sind nicht nur in der Prüfungsplanung und der Prüfungsdurchführung, sondern auch im Rahmen der abschließenden Gesamtdurchsicht einer Abschlussprüfung anzuwenden. Der Prüfer hat hierbei die einzelnen Prüfungsgebiete abschließend zu beurteilen und zu überprüfen, ob der Jahresabschluss insgesamt einen den wirtschaftlichen Verhältnissen des Unternehmens entsprechenden Gesamteindruck vermittelt.<sup>2</sup> Analytische Prüfungshandlungen unterstützen den Prüfer in dieser Phase der Prüfung bei der Erlangung von Erkenntnissen, die die bisherigen Prüfungsfeststellungen entweder bestätigen oder widerlegen.<sup>3</sup> Im Fall einer Bestätigung der bisherigen Feststellungen bieten analytische Prüfungshandlungen zum Ende der Prüfung eine zusätzliche Sicherheit, dass der Jahresabschluss frei von wesentlichen Fehlern ist.<sup>4</sup> Werden hingegen Ergebnisse erlangt, die gegen die Korrektheit der getroffenen Prüfungsfeststellungen sprechen, so signalisieren sie den Bedarf nach zusätzlichen Prüfungshandlungen.<sup>5</sup>

Dabei liegen die wesentlichen Ziele analytischer Prüfungshandlungen im Rahmen der abschließenden Gesamtdurchsicht in der *Berücksichtigung von Korrekturbuchungen oder Reklassifizierungen im Jahresabschluss*, die einen „anderen“ Jahresabschluss verursachen können, der *nachträglichen Identifikation bisher nicht erkannter Fehlerrisiken* und der *Beurteilung der Angemessenheit der erlangten Prüfungsnachweise*.

---

<sup>1</sup> Eine ausführliche Übersicht der möglichen analytischen Prüfungshandlungen im Rahmen der Prüfungsdurchführung findet sich etwa bei FITZSIMONS/LEVINE (1989), S. 19-21 sowie FLEMING (2004), S. 4.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 295-297.

<sup>3</sup> Vgl. PORTER/SIMON/HATHERLY (2003), S. 230-231.

<sup>4</sup> Vgl. IDW (2006), R 396; DEFLEISE/JAENICKE/O'REILLY/HIRSCH (1990), S. 593.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 297-298; GÄRTNER (1994), S. 91-92; BLOCHER/WILLINGHAM (1985), S. 7.

#### **2.1.2.4.1.2 Berücksichtigung von Korrekturbuchungen oder Reklassifizierungen im Jahresabschluss**

Analytische Prüfungshandlungen zum Ende einer Prüfung ermöglichen es dem Prüfer, eine abschließende Gesamtbeurteilung vorzunehmen, nachdem alle wesentlichen Korrekturbuchungen und Reklassifizierungen durchgeführt wurden.<sup>1</sup> Denn durch Änderungen in den Jahresabschlussdaten kann der Prüfer bei Anwendung von analytischen Prüfungshandlungen zu Ergebnissen gelangen, die von den Ergebnissen der Prüfungen in den anderen Phasen abweichen.<sup>2</sup> Daher sollte der Prüfer die wichtigen Prüfungsgebiete unter Berücksichtigung aller wesentlichen Korrekturbuchungen und Reklassifizierungen erneut einer Anwendung analytischer Prüfungshandlungen unterziehen.<sup>3</sup> Dabei ist es sinnvoll, die gleichen analytischen Methoden vorzunehmen, die bereits in der Phase der Planung eingesetzt wurden. Die nachträgliche Berechnung von Trends oder Kennzahlen kann das Vertrauen des Prüfers in die Korrektheit des Jahresabschlusses erhöhen, indem die Plausibilität des Jahresabschlusses als Ganzes bestätigt wird. Die Normenkonformität des Jahresabschlusses kann jedoch auch in Frage gestellt werden, da im Nachhinein Inkonsistenzen aufgedeckt werden, die in der Planungsphase aufgrund des „verfälschten“ Bilds unerkannt blieben. Darüber hinaus erscheint es zweckdienlich, im Rahmen der Prüfungsdurchführung eingesetzte analytische Prüfungshandlungen erneut zu kalkulieren. Dies ist insbesondere dort von Bedeutung, wo weitere Prüfungshandlungen aufgrund der Ergebnisse analytischer Prüfungshandlungen (auf Basis des „alten“ Datenmaterials) reduziert oder unterlassen wurden.

#### **2.1.2.4.1.3 Nachträgliche Identifikation bisher nicht erkannter Fehlerrisiken**

Der zweite wesentliche Vorteil einer abschließenden Gesamtdurchsicht zum Schluss einer Jahresabschlussprüfung liegt darin, dass durch eine prüffeldübergreifende Globalbetrachtung unter Berücksichtigung aller vorgenommenen Änderungen im Jahresabschluss noch nicht festgestellte Fehlerrisiken möglicherweise eher auffallen als durch detaillierte, aber schrittweise Untersuchungen einzelner Prüfungsgebiete.<sup>4</sup> PENNY be-

---

<sup>1</sup> Vgl. MCKEE (1989), S. 29.

<sup>2</sup> Vgl. GÄRTNER (1994), S. 94.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 140-141.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 141.

gründet diesen Vorteil damit, dass analytische Prüfungshandlungen in der abschließenden Gesamtdurchsicht „allow the auditor to see the big picture“.<sup>1</sup>

In diesem Kontext eröffnet sich zudem die Möglichkeit, dass der Prüfer alle festgestellten unerwarteten Abweichungen in einem Gesamtzusammenhang betrachtet. Dabei könnte der Prüfer etwa auf ein „Muster“ in den Jahresabschlussdaten stoßen, welches ein bisher unentdecktes Fehlerrisiko zu erkennen gibt. Denkbar wäre etwa, dass alle einzelnen Abweichungen für sich allein betrachtet auf kein größeres Risiko schließen lassen, jedoch die Betrachtung der einzelnen Abweichungen als ein Ganzes zu erkennen gibt, dass das Unternehmen durch weit gestreute bilanzgestalterische Maßnahmen versucht, positive Erfolgsbeiträge und Aktiva zu überbewerten oder negative Erfolgsbeiträge und Passiva zu unterbewerten.<sup>2</sup> Diese Erkenntnis könnte möglicherweise dazu führen, dass die Ergebnisse bisheriger analytischer Prüfungshandlungen in einem anderen Lichte zu betrachten wären, etwa dort, wo keine wesentlichen Abweichungen festgestellt wurden oder die Jahresabschlussdaten einem scheinbaren Trend folgen.

#### **2.1.2.4.1.4 Beurteilung der Angemessenheit der erlangten Prüfungsnachweise**

Neben der nachträglichen Aufdeckung von bisher nicht erkannten „Ungereimtheiten“ im Jahresabschluss dient der Einsatz von analytischen Prüfungshandlungen im Rahmen der abschließenden Gesamtdurchsicht auch der Beurteilung der Angemessenheit der erlangten Prüfungsnachweise.<sup>3</sup> Ein besonderes Augenmerk sollte der Prüfer auf die während der übrigen Prüfungsphasen identifizierten unerwarteten und wesentlichen Abweichungen richten und dabei noch einmal kritisch untersuchen, ob die durchgeführten Prüfungshandlungen ausreichend und angemessen sind.<sup>4</sup> Dabei sollte die Beurteilung der Angemessenheit weniger auf die Quantität als vielmehr auf die Qualität der erlangten Prüfungsnachweise ausgerichtet sein. So können etwa umfassende Checklisten und andere standardisierte Prüfungshandlungen auf einen ausreichenden Umfang eingeholter Prüfungsnachweise hindeuten, die Qualität der erlangten Prüfungsnachweise hingegen könnte unter anderem aufgrund des Einsatzes von unerfahrenen Prüfern im Nachhinein als unangemessen erscheinen. Eine abschließende Evaluierung sollte etwa in Bezug auf die Vertrauenswürdigkeit der Prüfungsnachweise erfolgen. Der Prüfer sollte sich etwa

---

<sup>1</sup> Vgl. PENNY (2003), S. 92.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 141.

<sup>3</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 58; MARTEN/QUICK/RUHNKE (2007), S. 297-298.

<sup>4</sup> Vgl. DEFLEISE/JAENICKE/O'REILLY/HIRSCH (1990), S. 593; MÜLLER (1996), S. 33 m.w.N.



die Frage stellen, inwiefern Prüfungsnachweise auf Informationen des Managements beruhen, die nicht durch eine unabhängige Quelle bestätigt wurden.<sup>1</sup>

#### 2.1.2.4.2 Prüfungsgegenstände analytischer Prüfungshandlungen in Rahmen der abschließenden Gesamtdurchsicht

Da während der abschließenden Gesamtdurchsicht sowohl die im Rahmen der Prüfungsplanung als auch die im Rahmen der Prüfungsdurchführung verwendeten analytischen Prüfungshandlungen erneut in Betracht zu ziehen sind, richten sich analytische Prüfungshandlungen in dieser Phase des Prüfungsprozesses auf hochaggregierte Daten einerseits und detaillierte Daten andererseits. Darüber hinaus ist es zudem möglich, weitere bisher noch nicht durchgeführte Verfahren analytischer Prüfungshandlungen einzusetzen.

Abschließend werden die Ziele und die Prüfungsgegenstände analytischer Prüfungshandlungen in allen drei Phasen einer Jahresabschlussprüfung graphisch wiedergegeben (siehe Abbildung 1).

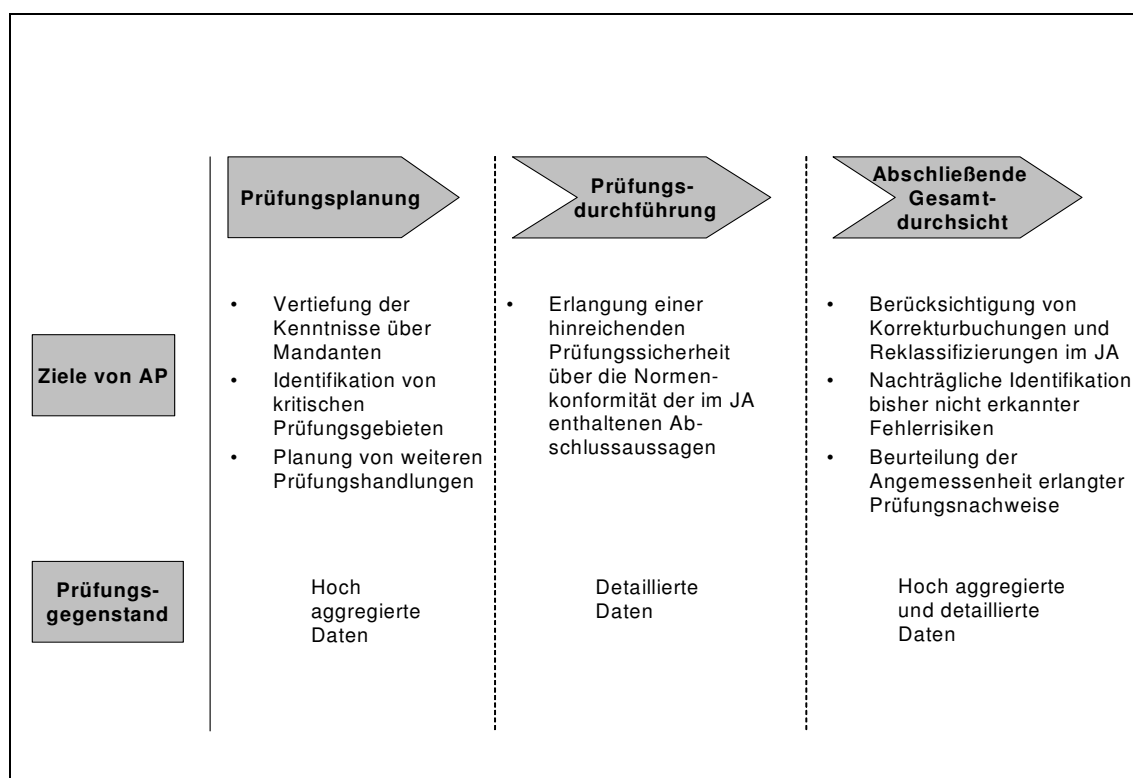


Abbildung 1: Ziele und Gegenstände analytischer Prüfungshandlungen in Abhängigkeit der Phase der Abschlussprüfung

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 141.

## 2.2 Verfahren analytischer Prüfungshandlungen

### 2.2.1 Grundlegende Vorbemerkungen

Die Effektivität und Effizienz von analytischen Prüfungshandlungen hängt im Wesentlichen davon ab, inwiefern analytische Prüfungshandlungen dazu in der Lage sind, dem Prüfer wesentliche Fehler in den zur Beurteilung stehenden Jahresabschlussdaten zu signalisieren. Dabei ist es von entscheidender Bedeutung, welchen Wert (sog. Erwartungswert) der Prüfer den zu prüfenden Jahresabschlussdaten gegenüberstellt.<sup>1</sup> BLOCHER/PATTERSON vergleichen diese prüferische Vorgehensweise mit einer ärztlichen Untersuchung, bei der die aktuellen Werte eines Patienten ebenfalls nicht angemessen interpretiert werden können, ohne zuvor einen Einblick in die medizinische Akte dieses Patienten genommen zu haben.<sup>2</sup>

Zur Bildung dieses Erwartungswerts stehen dem Prüfer unterschiedliche quantitative und qualitative Informationen zur Verfügung. In Abhängigkeit von den zur Bildung des Erwartungswerts herangezogenen Informationen lassen sich verschiedene Verfahren analytischer Prüfungshandlungen differenzieren. Die herangezogenen Informationen können dabei in drei Gruppen unterteilt werden, denen wiederum unterschiedliche Verfahren analytischer Prüfungshandlungen zugeordnet werden können. Dabei handelt es sich zunächst um *Informationen aus dem vergangenen Geschäftsjahr bzw. vergangenen Geschäftsjahren*. Verfahren, die auf vergangenheitsbezogenen Informationen basieren, können unterteilt werden in *einfache Vorjahresvergleiche*, *Trendanalysen* und *zeitbezogene Kennzahlenanalysen*. Darüber hinaus kann auf *Informationen aus dem aktuellen Geschäftsjahr* zurückgegriffen werden. Hier wird generell zwischen *Benchmark-Vergleichen* und *einfachen Rechenmodellen* differenziert. Da hierbei Informationen aus der Vergangenheit vollends außen vor bleiben, können diese Verfahren auch als *Ein-Perioden-Modelle* bezeichnet werden.<sup>3</sup> Schließlich existiert zudem die Möglichkeit, *Informationen aus der Vergangenheit und aus dem aktuellen Geschäftsjahr* heranzuziehen. Dieses geschieht regelmäßig bei dem Verfahren der *Regressionsanalyse*.

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 298, 300; ARENS/ELDER/BEASLEY (2005), S. 209.

<sup>2</sup> Vgl. BLOCHER/PATTERSON (1996), S. 53.

<sup>3</sup> In der Literatur werden häufig nur einfache Rechenmodelle als Ein-Perioden-Modelle bezeichnet (vgl. etwa MÜLLER (1996), S.38; GÄRTNER (1994), S. 107; MCKEE (1989), S. 114; BLOCHER/WILLINGHAM (1985), S. 101). Da aber auch bei den Benchmark-Vergleichen ausschließlich auf Daten zurückgegriffen wird, die der gleichen Periode wie der zu prüfende Jahresabschluss angehören, sollten auch diese den Ein-Perioden-Modellen zugeordnet werden.

## 2.2.2 Informationen aus der Vergangenheit

### 2.2.2.1 Einfacher Vorjahresvergleich

Eine sehr weit verbreite wenn auch sehr triviale Form der analytischen Prüfungshandlungen stellt der einfache Vorjahresvergleich dar.<sup>1</sup> Hierbei wird lediglich der zu prüfende Buchwert dem korrespondierenden Wert des Vorjahrs gegenübergestellt und auf Abweichungen untersucht. Sollte der Wert erheblich gegenüber dem Vorjahr gestiegen oder gesunken sein, wird der Prüfer dies zum Anlass nehmen, einen Fehler im Datenmaterial zu vermuten und weitere auf den tatsächlichen Grund für diese Abweichung gerichtete Untersuchungen vorzunehmen. Bei diesem im angloamerikanischen Sprachraum auch als *rule of thumb*<sup>2</sup> (Faustregel) bezeichneten Verfahren wird angenommen, dass der Wert des Vorjahrs eine geeignete erste Schätzung für den entsprechenden Wert des zu prüfenden Geschäftsjahrs darstellt.<sup>3</sup>

Der Prüfer kann sich bei einfachen Vorjahresvergleichen auf *absolute* und *relative* Veränderungsbeträge fokussieren. Der Prüfer wird sich dabei wegen der Möglichkeit wesentlicher Fehler stets auf große absolute Veränderungsbeträge konzentrieren. Allerdings sollte er auch weiterhin hohe relative Änderungen im Auge behalten, da sie einen Hinweis auf größere Probleme im Jahresabschluss geben können, auch wenn der absolute Veränderungsbetrag als unwesentlich aufgefasst werden kann.<sup>4</sup>

Die Schwierigkeit von einfachen Vorjahresvergleichen liegt in der Entscheidung des Prüfers, wann von einer signifikanten Abweichung im Vergleich zum Vorjahr ausgegangen werden kann, die die Vornahme von weiteren auf die Ursache der Abweichung gerichteten Untersuchungen unverzichtbar macht. Orientiert sich der Prüfer an *absoluten Beträgen*, so erscheint es angemessen, dass eine signifikante Abweichung in Abhängigkeit von der Wesentlichkeitsgrenze des entsprechenden Prüfungsauftrags (M) festgelegt wird.<sup>5</sup> Zur Berechnung der individuellen Wesentlichkeitsgrenze eines bestimmten Prüffelds ( $M_i$ ) in Abhängigkeit von dessen relativem Wert kann auf folgende Formel zurückgegriffen werden:

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 300.

<sup>2</sup> Vgl. etwa KINNEY/FELIX (1980), S. 101.

<sup>3</sup> Vgl. GAUNTT/GLEZEN (1997), S. 56.

<sup>4</sup> Vgl. MCKEE (1989), S. 59. Für eine gleichzeitige Untersuchung absoluter und relativer Veränderungsbeträge spricht sich KONRATH (2002), S. 119 aus.

<sup>5</sup> Vgl. MCKEE (1989), S. 59.

$$M_i = M \times \sqrt{\frac{BW_i}{BW}}$$

mit

$BW_i$  = Buchwert des betrachteten Prüffelds

$BW$  = Summe der Buchwerte aller Prüffelder

In Bezug auf *prozentuale Abweichungen* hat sich als sog. investigation rule<sup>1</sup> durchgesetzt, dass eine Abweichung immer dann als signifikant zu bezeichnen ist, wenn der zu prüfende Buchwert um mehr als 10% vom Vorjahreswert abweicht.<sup>2</sup>

Der Einsatz einfacher Vorjahresvergleiche ist besonders dann zu empfehlen, wenn die Geschäftstätigkeit des Mandanten sowie sein wirtschaftliches und rechtliches Umfeld vergleichbar zum Vorjahr sind.<sup>3</sup> In derartigen Fällen können Vorjahrsvergleiche als durchaus verlässliche Verfahren eingeschätzt werden, die sehr einfach und kostengünstig umgesetzt werden können.<sup>4</sup> Allerdings stoßen Vorjahresvergleiche schnell auf ihre Grenzen, wenn sich im zu prüfenden Geschäftsjahr erhebliche Änderungen im Vergleich zum Vorjahr ergeben haben.<sup>5</sup> Für derartige Fälle wäre es sogar empfehlenswert, auf einfache Vorjahresvergleiche zu verzichten, da sie die Effektivität und Effizienz der Jahresabschlussprüfung erheblich beeinträchtigen können. So wäre denkbar, dass ungeprüfte Buchwerte, die aufgrund von umweltbezogenen Einflüssen, wie etwa neuen Produkten oder Produktionsprozessen, Ein- bzw. Verkauf von Vermögensgegenständen des Anlagevermögens, verschärften Wettbewerbssituationen oder Beschaffungsproblemen, erheblich von den Vorjahreswerten abweichen, ohne dass ein tatsächlicher Fehler im Datenmaterial vorliegt. Sollte der Prüfer hier auf die Vornahme weiterer zeit- und kostenintensiver Einzelfallprüfungen bestehen, würden diese letztendlich lediglich die Normenkonformität des Datenmaterials feststellen. Dieses würde zwar nicht die Effektivität der Jahresabschlussprüfung gefährden, jedoch erhebliche Effizienzverluste zur Folge haben. Weitaus schwerwiegender wäre allerdings der Fall, in dem Veränderungen in der Geschäftstätigkeit oder im Unternehmensumfeld eine signifikante Abweichung mit sich bringen müssten, diese jedoch aufgrund wesentlicher Fehler im Datenmaterial kompensiert werden, so dass ein einfacher Vorjahresvergleich keine signifikante Abweichung

---

<sup>1</sup> Vgl. KINNEY/FELIX (1980), S. 101.

<sup>2</sup> Vgl. MCKEE (1989), S. 60; KINNEY/FELIX (1980), S. 101.

<sup>3</sup> Vgl. KINNEY/MCDANIEL (1996), S. 52.

<sup>4</sup> Vgl. KINNEY/FELIX (1980), S. 101.

<sup>5</sup> Vgl. GAUNTT/GLEZEN (1997), S. 56.

signalisiert. Würde der Prüfer folglich die Vornahme weiterer Prüfungshandlungen einschränken oder gar auslassen, so wären die Auswirkungen auf die Effektivität der Jahresabschlussprüfung fatal.

## **2.2.2.2 Einfache Trendanalysen**

### **2.2.2.2.1 Vorbemerkungen**

Dem im vorangegangenen Abschnitt beschriebenen einfachen Vorjahresvergleich kann insbesondere entgegengehalten werden, dass im zu prüfenden Geschäftsjahr vollzogene Änderungen im Umfang der Geschäftstätigkeit vollends unberücksichtigt bleiben.<sup>1</sup> Sollte sich das Unternehmen etwa seit Jahren in einer erheblichen Expansionsphase befinden, würden einfache Vorjahresvergleiche tendenziell auf die Überbewertung des zu prüfenden Buchwerts hindeuten. Sollte sich der Prüfer deshalb für die Ausweitung der Prüfungshandlungen entscheiden, wäre mit erheblichen Effizienzeinbußen zu rechnen.

Eine Möglichkeit, dieser Problematik zu entgehen, ist in der Anwendung der sog. *einfachen Trendanalyse* (im Folgenden als Trendanalyse bezeichnet) zu sehen.<sup>2</sup> Diese ist im Gegensatz zum einfachen Vorjahresvergleich nicht nur auf einen einzelnen Wert, den Vorjahreswert, ausgerichtet. Vielmehr finden hier Daten einer größeren Zeitspanne Berücksichtigung. Dieses hat zum Vorteil, dass zur Beurteilung der Plausibilität eines zu prüfenden Werts insbesondere die Entwicklung des Unternehmens im Zeitablauf als Beurteilungskriterium herangezogen wird. Der Prüfer ist bei der Entwicklung eines Erwartungswerts somit in der Lage, einen Trend in den Daten zu berücksichtigen.<sup>3</sup> Ferner ermöglicht es dem Prüfer zu erkennen, ob der Vorjahreswert möglicherweise einen Ausreißer darstellt und somit in der Gesamtbeurteilung nicht repräsentativ ist.<sup>4</sup> In diesen Fällen würde ein Vorjahresvergleich zu verzerrten Erwartungswerten führen.

Bei der Anwendung von Trendanalysen stellt sich die Frage, ob der Prüfer auf Jahres-, Quartals- oder Monatswerte zurückgreifen soll. Wählt der Prüfer geringe Zeitabstände zwischen den einzelnen Beobachtungswerten (etwa Monats- oder Quartalswerte), ste-

---

<sup>1</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 211.

<sup>2</sup> Hiervon abzugrenzen ist die sog. *kennzahlenbasierte Trendanalyse*, welche im Gegensatz zur einfachen Trendanalyse nicht die Entwicklung einer absoluten Zahl (hier der zu prüfende Buchwert) betrachtet, sondern die Entwicklung einer jahresabschlussbasierten relativen Zahl. Siehe hierzu ausführlich Kap. 2.2.2.3.

<sup>3</sup> Vgl. GAUNTT/GLEZEN (1997), S. 56.

<sup>4</sup> Vgl. MCKEE (1989), S. 60.

hen ihm generell mehr Datensätze zur Verfügung, was in der Regel zu einer höheren Genauigkeit der zu entwickelnden Erwartungswerte führt. Allerdings stehen dem Prüfer nicht immer Quartals- oder Monatswerte zur Verfügung. Falls der Prüfer auf unterjährige Daten zurückgreifen kann, ist zu berücksichtigen, dass es sich regelmäßig um ungeprüfte Werte handelt. Potenziell vorhandene Fehler im Datenmaterial, etwa monatliche cutoff errors, würden folglich die Beurteilung des Prüfers negativ beeinflussen.<sup>1</sup> Zudem sollten die verwendeten Daten durch unterjährige Stabilität gekennzeichnet sein und keine saisonalen Schwankungen aufweisen.

Als wesentlicher Kritikpunkt an der Trendanalyse lässt sich festhalten, dass die Zeit als einziger Einflussfaktor für die Entwicklung der untersuchten Jahresabschlussposition betrachtet wird, obgleich sie keinen eigenständigen Erklärungswert für den Verlauf einer Jahresabschlussposition inne hat. Vielmehr unterstellt die Trendanalyse, dass alle auf die zu prognostizierende Jahresabschlussgröße einwirkenden Faktoren im Zeitablauf konstant bleiben. Aufgrund dieser sehr stark vereinfachenden Annahme droht die sog. Extrapolationsfalle, da die in der Realität existierenden wechselseitigen Beziehungen zwischen den einzelnen Jahresabschlussgrößen außen vor bleiben.<sup>2</sup>

#### **2.2.2.2.2 Vorgehensweise des Prüfers bei Trendanalysen**

Bei der Trendanalyse stellt der Prüfer den ausgewiesenen Buchwert den Werten der vorangegangenen Jahre gegenüber und beurteilt, ob der Buchwert in Anbetracht des sich im Zeitablaufs verzeichneten Trends bzw. Musters plausibel erscheint.<sup>3</sup> Falls der Buchwert keine auffällige Abweichung vom Trend erkennen lässt, geht der Prüfer davon aus, dass keine wesentlichen Fehler im Datenmaterial existieren.<sup>4</sup> Da Jahresabschlusspositionen aber auch dann wesentliche Fehler beinhalten können, wenn keine Abweichung vom scheinbaren Trend zu erkennen ist, birgt die Anwendung der Trendanalyse Gefahren für die Effektivität der Jahresabschlussprüfung. Umgekehrt können scheinbare Abweichungen vom Trend auf Bedingungen zurückgeführt werden, die keinen Fehler im Jahresabschluss darstellen. In diesen Fällen würde die Trendanalyse die Effizienz der Prüfung erheblich beeinträchtigen.

---

<sup>1</sup> Vgl. KINNEY/FELIX (1980), S. 101; GÄRTNER (1994), S. 112.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 304.

<sup>3</sup> Vgl. LEE/COLBERT (1997), S. 393.

<sup>4</sup> Vgl. GÄRTNER (1994), S. 113.

Dennoch sollte die Trendanalyse nicht grundsätzlich abgelehnt werden. Ihr wesentlicher Vorteil ist in der einfachen Handhabung und dem vergleichsweise geringen Zeitaufwand zu sehen. Sollte das zu prüfende Unternehmen im abgelaufenen Geschäftsjahr keine wesentlichen Änderungen in der Geschäftstätigkeit und im unternehmerischen Umfeld erfahren haben, die eine wesentliche Abweichung des zu prüfenden Buchwerts vom Trend begründen, scheint die Trendanalyse durchaus angemessen. Dies gilt insbesondere für die Phase der Prüfungsplanung, um erste, noch sehr undifferenzierte Plausibilitätsbeurteilungen vornehmen zu können.<sup>1</sup>

#### **2.2.2.2.3 Methoden der Trendanalyse**

##### **2.2.2.2.3.1 Übersicht**

Bei der Trendanalyse können grundsätzlich fünf verschiedene Methoden differenziert werden, die sich in ihrer Komplexität voneinander unterscheiden. Es handelt sich hierbei um

- die graphische Methode,
- die Methode der durchschnittlichen Änderung,
- die Methode der gewichteten durchschnittlichen Änderung,
- die Methode des gleitenden Durchschnitts und
- die Methode der kleinsten Quadrate.

Grundsätzlich gilt, dass mit steigender Komplexität der Methode gleichzeitig die Verlässlichkeit des generierten Erwartungswerts zunimmt.<sup>2</sup>

##### **2.2.2.2.3.2 Graphische Methode**

Die einfachste Methode stellt die *graphische Methode* dar. Hierbei trägt der Prüfer die Werte des zur Beurteilung stehenden Jahresabschlusspostens zu unterschiedlichen Zeitpunkten in ein Streudiagramm ein. Zur Entwicklung des Erwartungswerts zeichnet der Prüfer zunächst freihändig eine Trendlinie in das Diagramm ein und verlängert diese

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 114.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 51.

anschließend bis zu dem Zeitpunkt, für den der ungeprüfte Buchwert beurteilt werden soll.<sup>1</sup> Falls bei der Eintragung der Trendlinie ein Datenpunkt auffällig von den anderen Punkten abweicht, kann der Prüfer diesen als einen sog. Ausreißer betrachten und ignorieren.<sup>2</sup> Allerdings sollte sich der Prüfer darüber vergewissern, was die Umstände dieses Ausreißers waren, und dass diese Umstände nicht auch im zu prüfenden Geschäftsjahr vorliegen.<sup>3</sup>

Eine wesentliche Schwierigkeit bei der Anwendung der graphischen Methode stellt die Frage dar, wann von einer signifikanten Abweichung des Buchwerts von dem gebildeten Erwartungswert auszugehen ist. Hierzu stehen dem Prüfer zwei Möglichkeiten zur Verfügung. Einerseits kann er den Abweichungsbetrag und diesen vor dem Hintergrund der gegebenen Wesentlichkeitsgrenzen bewerten.<sup>4</sup> Sollte sich ergeben, dass der Betrag als unwesentlich aufzufassen ist, so kann dies als eine erste Einschätzung über die Normenkonformität der Jahresabschlussposition gewertet werden.<sup>5</sup>

Andererseits kann der Prüfer zur Einschätzung der Signifikanz von Abweichungen eine obere und eine untere Grenzlinie in das Streudiagramm einzeichnen. Die Grenzlinien sollten äquidistant zur Trendlinie verlaufen und nach Möglichkeit alle Datenpunkte, ausgenommen der Ausreißer, eingrenzen.<sup>6</sup> Die Grenzlinien können dabei als Konfidenzintervall verstanden werden, in dem sich der tatsächliche Buchwert mit einer bestimmten Wahrscheinlichkeit aufhält.<sup>7</sup> Falls der Buchwert des zu prüfenden Geschäftsjahrs innerhalb dieser Grenzlinien fällt, kann dies als erster Hinweis für die Normenkonformität des ausgewiesenen Buchwerts aufgefasst werden.<sup>8</sup>

Als äußerst unpräzise erweist sich die graphische Methode allerdings, falls das Konfidenzintervall aufgrund einer starken Streuung der Daten sehr groß ausfällt.<sup>1</sup> Um eine Einschätzung der Güte der graphischen Methode vornehmen zu können, sollte der Prüfer den Abstand zwischen der Trend- und den Grenzlinien in Relation zur prüffeldspezi-

---

<sup>1</sup> Vgl. im folgenden MARTEN/QUICK/RUHNKE (2007), S. 303-305.

<sup>2</sup> Zur Identifikation von Ausreißern werden in der Literatur unterschiedliche Vorgehensweisen vorgeschlagen. Die Vorschläge reichen von einfachen subjektiven Entscheidungen oder der Festsetzung von „Wahrscheinlichkeitsgrenzen“ um den Mittelwert, um jeden diese Grenzen überschreitenden Wert als Ausreißer zu identifizieren, bis hin zu objektiven Verfahren der automatischen Extremwertbereinigung. Vgl. hierzu ausführlich HÜTTNER (1986), S. 29-30 i.V.m. 307-310.

<sup>3</sup> Vgl. MCKEE (1989), S. 67.

<sup>4</sup> Der Abweichungsbetrag ergibt sich dabei als Differenz zwischen dem Wert, der aus der Verlängerung der Trendlinie bis zum zu prüfenden Geschäftsjahr resultiert, und dem ungeprüften Buchwert.

<sup>5</sup> Vgl. MCKEE (1989), S. 65.

<sup>6</sup> Vgl. MCKEE (1989), S. 65.

<sup>7</sup> Vgl. GÄRTNER (1994), S. 116.

<sup>8</sup> Vgl. MCKEE (1989), S. 65.



fischen Wesentlichkeitsgrenze setzen. Übersteigt der Abstand dabei die prüffeldspezifische Wesentlichkeitsgrenze, ist das Modell lediglich zur Erlangung von schwachen Prüfungsnachweisen in der Lage.<sup>2</sup>

Ein wesentlicher Kritikpunkt ist des Weiteren in der hohen Subjektivität der graphischen Methode zu sehen. Da die Trendlinie freihändig von dem Prüfer in das Streudiagramm eingezeichnet wird, ist anzunehmen, dass diese Linie in Abhängigkeit von der jeweiligen Person unterschiedlich ausfallen würde.<sup>3</sup> Auch der Einfluss von Ausreißern sowohl auf die Trendlinie als auf das Konfidenzintervall wird intersubjektiv möglicherweise nicht identisch ausfallen.

Eine Möglichkeit, diesen subjektiven Einflüssen entgegenzuwirken, ist in der Verwendung einfacher mathematischer Berechnungen des Trends in den Daten zu sehen, wie sie in den folgenden Methoden zu finden sind.

#### **2.2.2.2.3.3 Methode der durchschnittlichen Änderung**

Diese Methode basiert auf der Annahme, dass die durchschnittlichen Änderungen in dem betrachteten Zeitraum der Änderung der zu prüfenden Periode zum Vorjahr entspricht. Die durchschnittliche Änderung kann dabei sowohl als absoluter als auch als relativer Betrag errechnet werden, wobei jeweils unterschiedliche Erwartungswerte erreicht werden.<sup>4</sup> Im Folgenden soll beispielhaft gezeigt werden, wie sich die Ermittlung des Erwartungswerts unter Verwendung absoluter Werte ergibt:

$$BW_{\text{Soll}} = BW_n + (BW_1 - BW_n)/N$$

mit

$BW_{\text{Soll}}$  = zu prognostizierender Buchwert

$BW_n$  = Buchwert des Vorjahrs

$N$  = Anzahl der betrachteten Vorjahre

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 52.

<sup>2</sup> Ähnlich argumentiert MCKEE (1989), S. 68-69.

<sup>3</sup> Vgl. GÄRTNER (1994), S. 122.

<sup>4</sup> Vgl. MCKEE (1989), S. 69-70.

#### **2.2.2.2.3.4 Methode der gewichteten durchschnittlichen Änderung**

Die Methode der durchschnittlichen Änderungen ist dadurch gekennzeichnet, dass sie alle Änderungen von einem Zeitpunkt zum darauf folgenden gleichermaßen in die Bildung des Erwartungswerts einfließen lässt. Ist der Prüfer allerdings der Ansicht, dass einige Änderungen, etwa die in der jüngeren Vergangenheit, repräsentativer für die aktuelle Entwicklung sind, sollte er dieses durch eine entsprechende Gewichtung der Änderungen berücksichtigen. Eine einfache Vorgehensweise zur Umsetzung der Methode der gewichteten durchschnittlichen Änderungen wäre etwa denkbar, indem die Werte der Perioden mit ihrer Periodenzahl multipliziert werden. D.h. der Wert der ersten Periode wird mit eins multipliziert, der Wert der zweiten Periode mit 2, usw. Anschließend werden die ermittelten Produkte summiert und durch die Quersumme der Zahlenreihe 1 – n dividiert. Zur Ermittlung des Erwartungswerts der aktuellen Periode wird dieser Quotient zum Vorjahreswert hinzugerechnet.<sup>1</sup> Da die Änderungen hier unterschiedlich gewichtet werden, ergibt sich zwangsläufig ein anderer Wert als nach der Methode der durchschnittlichen Änderung.

#### **2.2.2.2.3.5 Methode des gleitenden Durchschnitts**

Eine weitere Möglichkeit, den Werten der jüngeren Vergangenheit einen höheren Einfluss zuzuordnen ist in der Methode des gleitenden Durchschnitts zu sehen. Vergleichbar zur vorhergehenden Methode wird prinzipiell auch hier den zeitlich näher liegenden Werten eine höhere Gewichtung zugeordnet. Allerdings werden hier nicht alle Datensätze berücksichtigt, sondern nur diejenigen, die vom Prüfer als Ausprägung des aktuell zu verzeichnenden saisonalen Schwankungsmusters aufgefasst werden.<sup>2</sup> Diese Methode zeichnet sich dadurch aus, dass “the average ‘moves’ forward and is updated as new data are added.”<sup>3</sup>

Abschließend soll noch einmal darauf hingewiesen werden, dass auch die durch die unterschiedlichen Konzeptionen der vorgestellten Methoden bedingten Variationen in den potenziellen Erwartungswerten nicht vollends dazu in der Lage sind, prüferabhängige Differenzen zu überwinden. Insofern sind auch diese auf einfachen mathematischen Berechnungen basierenden Methoden ähnlich wie die graphische Methode durch eine ge-

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 57-58.

<sup>2</sup> Vgl. MCKEE (1989), S. 71-72.

<sup>3</sup> BLOCHER/WILLINGHAM (1985), S. 57-58.

wisse Subjektivität gekennzeichnet. Allerdings weisen sie gegenüber der graphischen Methode den erheblichen Vorteil auf, dass die Vorgehensweise zur Ermittlung des Erwartungswerts aufgrund des vorhandenen Formalismus objektiv besser nachvollzogen werden kann und den interpersonellen Ermessensräumen ein gewisser Einhalt geboten wird.

#### **2.2.2.2.3.6 Methode der kleinsten Quadrate**

Bei der Methode der kleinsten Quadrate werden im Gegensatz zu den beiden vorherigen Methoden die Beobachtungswerte nicht unterschiedlich gewichtet, vielmehr gehen alle Werte mit der gleichen Gewichtung in die Berechnung ein. Da es sich bei der Trendanalyse um eine Sonderform der *Regressionsanalyse* handelt, entspricht die Trendberechnung im Wesentlichen der Vorgehensweise zur Ermittlung einer Regressionsfunktion.<sup>1</sup> Die Trendberechnung mittels der Methode der kleinsten Quadrate entspricht einer Regression auf die Zeit, womit sich bei der Ermittlung eines linearen Trends gegenüber einer linearen Regressionsfunktion<sup>2</sup> erhebliche Vereinfachungen ergeben.<sup>3</sup> Für eine lineare Trendfunktion lautet die Notation

$$Y = a + bt$$

mit

Y = abhängige Variable

t = erklärende Variable (hier die Zeit)

a = absolutes Glied

b = Koeffizient.

Durch die Vornahme der partiellen Ableitungen nach a und b sowie deren Nullsetzung ergeben sich die beiden sog. Normalgleichungen, anhand derer sich die beiden Parameter a und b wie folgt bestimmen lassen:

---

<sup>1</sup> Vgl. HÜTTNER (1986), S. 17.

<sup>2</sup> Siehe auch die Ermittlung der Regressionsfunktion in Kap. 2.2.4.2.2.

<sup>3</sup> Vgl. KREIB/NEUHAUS (2006), S. 9-10.

$$a = \frac{\frac{1}{n} \sum_{i=1}^n t_i y_i - \frac{1}{n} \sum_{i=1}^n t_i \cdot \frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{n} \sum_{i=1}^n t_i^2 - \left( \frac{1}{n} \sum_{i=1}^n t_i \right)^2},$$

$$b = \frac{1}{n} \sum_{i=1}^n y_i - a \cdot \frac{1}{n} \sum_{i=1}^n t_i.$$

Für  $t = 1, \dots, n$  werden nunmehr nicht die jeweiligen Jahreszahlen eingesetzt, vielmehr werden Kennziffern (im Folgenden mit  $t^1$  bezeichnet) verwendet, die so zu wählen sind, dass ihre Summe null ergibt. Folglich vereinfachen sich die Nullgleichungen und die Lösungen. Für Letztere ergeben sich

$$a = \frac{\sum_{i=1}^n t_i^1 y_i}{\sum_{i=1}^n t_i^1{}^2} \quad \text{sowie}$$

$$b = \frac{\sum_{i=1}^n y_i}{n}.$$

Wurden die beiden Parameter  $a$  und  $b$  ermittelt, liegt dem Prüfer eine Trendfunktion vor, die das Verhalten der zu prüfenden Größe in Abhängigkeit von der Zeit wiedergibt. Der Prüfer ist anschließend in der Lage, für den im zu prüfenden Geschäftsjahr vorliegenden Buchwert einen Erwartungswert zu ermitteln.

---

<sup>1</sup> Vgl. HÜTTNER (1986), S. 18-19.

### 2.2.2.3 Zeitbezogene Kennzahlenanalysen<sup>1</sup>

#### 2.2.2.3.1 Vorüberlegungen

Der in der Literatur am häufigsten hervorgebrachte Kritikpunkt an der einfachen Trendanalyse und dem einfachen Vorjahresvergleich liegt in der Verwendung von einzelnen Jahresabschlussgrößen, also absoluten Zahlen.<sup>2</sup> Änderungen in einzelnen Jahresabschlusspositionen sind jedoch nur selten auf Fehler im Datenmaterial zurückzuführen, in der Regel beruhen Änderungen häufiger auf Modifikationen in der Geschäftstätigkeit oder Unternehmensumwelt.<sup>3</sup> Daher erscheint es wenig geeignet, die Plausibilität von einzelnen Jahresabschlussgrößen allein anhand ihrer Entwicklung über die Zeit zu beurteilen. Vielmehr sollte der Prüfer versuchen, die wesentlichen Einflussfaktoren einer Jahresabschlussposition zu identifizieren und diese bzw. deren parallele Entwicklungen ebenfalls in seine Urteilsfindung einfließen zu lassen. Eine Möglichkeit, derartige Abhängigkeiten im Jahresabschluss zu berücksichtigen, liegt in der Verwendung der Kennzahlen, denn „the use of ratios exploits that account balances may be correlated due to both the double-entry nature of accounting and multiple accounting effects of management decision strategies.“<sup>4</sup> Insofern erscheint die Verwendung von Kennzahlen, bei denen die gegenseitige Beeinflussung unterschiedlicher Jahresabschlusspositionen bereits antizipiert wird, wesentlich besser für die Beurteilung der Plausibilität von Jahresabschlusspositionen geeignet als die Verwendung von absoluten Zahlen.<sup>5</sup>

#### 2.2.2.3.2 Begriffsbestimmung

Kennzahlen sind allgemein als hochverdichtete Maßgrößen zu verstehen, die in konzentrierter Form über einen bestimmten betriebswirtschaftlichen Sachverhalt infor-

---

<sup>1</sup> Neben zeitbezogenen Kennzahlenanalysen (sog. *time series approach*) existieren zudem branchenbezogene bzw. betriebsübergreifende Kennzahlenanalysen (sog. *cross sectional approach*). Während es sich bei Ersterer um eine Mehrperiodenbetrachtung handelt, stellt der Prüfer bei Letzterer nicht auf die zeitliche Entwicklung einer Kennzahl ab. Vielmehr wird hier die gleiche Kennzahl für eine identische Periode betrachtet. Dabei kann die Kennzahl eines Unternehmens mit einem Branchendurchschnittswert, mit der gleichen Kennzahl eines anderen Unternehmens bzw. Betriebs verglichen werden. (vgl. MCKEE (1989), S. 76-79). Da bei derartigen Kennzahlenanalysen allerdings auf Informationen zurückgegriffen wird, die aus dem aktuellen Geschäftsjahr stammen (vgl. hierzu ausführlich Kap. 2.2.3), werden diese hier nicht weiter betrachtet. Branchenbezogene bzw. betriebsübergreifende Kennzahlenanalysen bilden vielmehr den Gegenstand der Kap. 2.2.3.1.3 sowie 2.2.3.1.4.

<sup>2</sup> Vgl. etwa MARTEN/QUICK/RUHNKE (2007), S. 301; MÜLLER (1996), S. 40; GÄRTNER (1994), S. 122; BLOCHER/PATTERSON (1996), S. 54; MCKEE (1989), S. 75.

<sup>3</sup> Vgl. MÜLLER (1996), S. 40-41.

<sup>4</sup> Vgl. ARRINGTON/HILLISON/ICERMAN (1983), S. 158.

<sup>5</sup> Vgl. SPERL (1978), S. 218.

mieren.<sup>1</sup> Sie sollen die Menge der Jahresabschlussdaten auf wenige aber zentrale Größen verdichten und dabei auf relativ einfache Art und Weise komplexe Organisationsstrukturen und Prozesse abbilden.<sup>2</sup> In der Literatur werden häufig auch absolute Zahlen zu den Kennzahlen gezählt.<sup>3</sup> Da man sich bei der Betrachtung von absoluten Zahlen im Bereich des einfachen Vorjahresvergleichs<sup>4</sup> oder der einfachen Trendanalyse<sup>5</sup> befindet, werden im vorliegenden Kapitel ausschließlich relative Zahlen als Kennzahlen verstanden.

Kennzahlen in Form relativer Zahlen versuchen zwei oder mehrere Jahresabschlussgrößen, die in einem bestimmten Zusammenhang stehen, zueinander ins Verhältnis zu setzen. In Abhängigkeit davon, welche Beziehung zwischen den Jahresabschlussgrößen im Zähler und Nenner besteht, können Kennzahlen in *Gliederungszahlen*, *Beziehungszahlen* und *Messzahlen* klassifiziert werden.<sup>6</sup> Bei Gliederungszahlen erfolgt eine Gegenüberstellung einer Teilgröße und ihrer entsprechenden Gesamtgröße.<sup>7</sup> Beispielsweise kann der Personalaufwand dem Gesamtaufwand gegenübergestellt werden oder das Umlaufvermögen dem Gesamtvermögen. Auf diesem Wege kann das relative Gewicht einer Teilmenge im Verhältnis zum Gesamten aufgezeigt werden.<sup>8</sup> Beziehungszahlen hingegen setzten artverschiedene Größen in Relation zueinander.<sup>9</sup> Die in den Quotienten einfließenden Größen sollten dabei entweder in einer kausalen Beziehung zueinander stehen, wie etwa bei der Gegenüberstellung von Forderungen und Umsatzerlösen, oder in einer rein zuordnungstechnischen Weise, wie es etwa bei den Deckungsgraden (z.B. Anlagevermögen zu Eigenkapital) nach dem Kriterium der Fristenkongruenz erfolgt.<sup>10</sup> Messzahlen dagegen stellen die gleiche Größe zu unterschiedlichen Zeitpunkten oder aus unterschiedlichen Zeiträumen gegenüber. Mit ihnen lassen sich Änderungen oder Entwicklungen einzelner Größen im Zeitablauf darstellen.<sup>11</sup> So können zum Beispiel die Umsätze des zu prüfenden Geschäftsjahrs in Relation zu den Umsätzen des Vorjahrs gesetzt werden.

---

<sup>1</sup> Vgl. KÜTING/WEBER (2006), S. 51.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 446.

<sup>3</sup> Vgl. BAETGE/KIRSCH/THIELE (2004), S. 147; COENENBERG (2005), S. 971; LACHNIT (2004), S. 39; KÜTING/WEBER (2006), S. 51.

<sup>4</sup> Vgl. ausführlich Kap. 2.2.2.1.

<sup>5</sup> Vgl. ausführlich Kap. 2.2.2.2.

<sup>6</sup> Vgl. LÜCK (1998), S. 437-438.

<sup>7</sup> Vgl. LACHNIT (2004), S. 39.

<sup>8</sup> Vgl. COENENBERG (2005), S. 971.

<sup>9</sup> Vgl. LACHNIT (2004), S. 39.

<sup>10</sup> Vgl. COENENBERG (2005), S. 972.

<sup>11</sup> Vgl. KÜTING/WEBER (2006), S. 54.

Zur Beurteilung der Plausibilität von Jahresabschlussdaten eignet sich insbesondere der Einsatz von Gliederungs- und Beziehungszahlen, da sie es ermöglichen, die funktionalen Zusammenhänge zwischen den einzelnen Jahresabschlussgrößen adäquat abzubilden. Die Verwendung von Messzahlen ist hingegen eher dem Bereich der Unternehmensanalyse zuzuordnen, da Veränderungen von einzelnen Jahresabschlussgrößen im Zeitablauf eher auf Entwicklungen und Modifikationen der Geschäftstätigkeit zurückgeführt werden können als auf Fehler in den Jahresabschlussdaten.<sup>1</sup>

#### **2.2.2.3.3      Arten von zeitbezogenen Kennzahlenanalysen**

Zeitbezogene Kennzahlenanalysen stellen die gleiche Kennzahl *eines Unternehmens* aus *unterschiedlichen Geschäftsjahren* gegenüber.<sup>2</sup> Dabei kann sich der Vergleich auf lediglich *zwei Perioden* beziehen. Hierbei käme die prüferische Vorgehensweise dem einfachen Vorjahrvergleich sehr nahe,<sup>3</sup> allerdings mit dem Unterschied, dass der Vergleich hier auf Basis von Kennzahlen und nicht von absoluten Zahlen erfolgt. In diesem Zusammenhang könnte der Kennzahlenvergleich über zwei Perioden auch als *kennzahlenbasierter Vorjahresvergleich* bezeichnet werden. Der Zeitvergleich kann aber auch über eine *Vielzahl von Perioden* erfolgen.<sup>4</sup> In diesem Fall inspiziert der Prüfer die Werte der einbezogenen Perioden auf Entwicklungen oder Muster im Zeitablauf und zieht diese als Basis für die Beurteilung der Plausibilität der zu prüfenden Jahresabschlussgröße heran. Diese Möglichkeit der Kennzahlenanalyse kann auch als *kennzahlenbasierte Trendanalyse* bezeichnet werden.

Im Folgenden sollen unterschiedliche Kennzahlen vorgestellt werden, auf die sowohl im Rahmen eines kennzahlenbasierten Vorjahresvergleichs als auch bei einer kennzahlenbasierten Trendanalyse zurückgegriffen werden kann.

---

<sup>1</sup> Vgl. SPERL (1978), S. 218-219.

<sup>2</sup> Vgl. LEE/COLBERT (1997), S. 394.

<sup>3</sup> Vgl. hierzu Kap. 2.2.2.1.

<sup>4</sup> Vgl. MCKEE (1989), S. 77.

### 2.2.2.3.4 Geeignete Kennzahlen zur Beurteilung der Plausibilität von Jahresabschlussdaten

#### 2.2.2.3.4.1 Überblick

Bei analytischen Prüfungshandlungen in Form von zeitbezogenen Kennzahlenanalysen kommt es entscheidend darauf an, dass die zueinander in Relation gesetzten Jahresabschlussgrößen in einem erklärbaren Zusammenhang zueinander stehen und ihre Entwicklungen unter normalen Umständen im gleichen Verhältnis verlaufen, denn ausschließlich bei im Zeitablauf relativ konstanten Kennzahlen kann im Falle einer unerwarteten Abweichung von einem erhöhten Fehlerrisiko ausgegangen werden.<sup>1</sup> In der Literatur existiert eine unüberschaubare Menge von Kennzahlen, die diese Kriterien erfüllen. Da der Prüfer nach IDW PS 312.9 letztendlich eigenverantwortlich für die Auswahl einer geeigneten analytischen Prüfungshandlung ist, sollte er die jeweils für den Prüfungszweck relevanteste und aussagekräftigste Kennzahl verwenden.<sup>2</sup>

Die in der Literatur am häufigsten anzutreffende Systematisierung geeigneter Kennzahlen (common ratios) erfolgt nach *finanzwirtschaftlichen Aspekten*.<sup>3</sup> Im Allgemeinen wird dabei zwischen

- Rentabilitätskennzahlen<sup>4</sup>,
- Liquiditätskennzahlen<sup>5</sup> sowie
- Umschlagskennzahlen<sup>1</sup>

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 41.

<sup>2</sup> Vgl. MCKEE (1982), S. 40.

<sup>3</sup> Vgl. etwa GÄRTNER (1994), S. 130-131; KNECHEL (2001), S. 276-279; KONRATH (2002), S. 120-121; ARENS/ELDER/BEASLEY (2005), S. 215-216; FITZSIMONS/LEVINE (1989) S. 10-17.

<sup>4</sup> Bei Rentabilitätskennzahlen wird eine Ergebnisgröße ins Verhältnis zu einer vermuteten Einflussgröße gesetzt (siehe ausführlich BAETGE/KIRSCH/THIELE (2004), S. 349-350). Häufig herangezogene Rentabilitätskennzahlen im Rahmen von analytischen Prüfungshandlungen sind beispielsweise die Gesamtkapital- oder die Eigenkapitalrentabilität (vgl. etwa ARENS/ELDER/BEASLEY (2005), S. 219 sowie KNECHEL (2001), S. 277). Da diese beiden Kennzahlen allerdings auf hochverdichteten Daten basieren, sollte sich ihr Einsatz überwiegend auf die Phase der Prüfungsplanung beschränken. Eine weitere Rentabilitätskennzahl ist die Bruttogewinnmarge, welche ausführlich in Kap.2.2.2.3.4.3.1 behandelt wird.

<sup>5</sup> Die Ermittlung von Liquiditätskennzahlen erfolgt durch die Gegenüberstellung von Posten des Umlaufvermögens zu den kurzfristigen Verbindlichkeiten. In Abhängigkeit von den jeweils einbezogenen Posten des Umlaufvermögens können unterschiedliche Kennzahlen berechnet werden, die auch als Liquiditätsgrade bezeichnet werden (siehe ausführlich BAETGE/KIRSCH/THIELE (2004), S. 263). In der Prüfungsliteratur wird zumeist auf den 2. und 3. Liquiditätsgrad verwiesen, welche im angloamerikanischen Sprachraum auch als *quick ratio* bzw. *current ratio* bezeichnet werden (vgl. etwa KNECHEL (2001), S. 277 oder KONRATH (1999), S. 121). Da diese beiden Kennzahlen jedoch aus einer Mehrzahl hochverdichteter Daten bestehen, sollten sie ausschließlich im Rahmen der Prüfungsplanung eingesetzt werden.



differenziert. Da diese Klassifizierung jedoch eher an die Jahresabschlussanalyse im Zusammenhang mit der Beurteilung der Geschäftstätigkeit eines Unternehmens und weniger an die einzelnen prüfungsrelevanten Bereiche im Rahmen der Jahresabschlussprüfung erinnert,<sup>2</sup> wird im Folgenden auf eine Systematisierung nach finanzwirtschaftlichen Gesichtspunkten zugunsten einer verstärkt prüffeldspezifischen Darstellung von Kennzahlen verzichtet. Dabei sollen ausschließlich solche Prüfungsbereiche berücksichtigt werden, die über klare und über die Zeit stabile Ursache-Wirkungszusammenhänge zu anderen Bereichen verfügen. Derartige kausale Beziehungen bestehen beispielsweise im Bereich

- des *Sachanlagevermögens*,
- der *Forderungen*,
- der *Vorräte*,
- der *Verbindlichkeiten und Rückstellungen* sowie
- der *Gewinn- und Verlustrechnung*.

Ungeeignet für den Einsatz von Kennzahlenanalysen ist hingegen der Bereich der *liquiden Mittel*, da die Zahlungsströme eine Vielzahl von anderen Konten betreffen können und keine eindeutige „eins-zu-eins“-Beziehung zu einem anderen Prüfungsbereich vorliegt.<sup>3</sup>

#### **2.2.2.3.4.2 Bilanz**

##### **2.2.2.3.4.2.1 Sachanlagevermögen**

Obwohl sich der Bereich des Sachanlagevermögens aufgrund der relativen Unregelmäßigkeit anfallender Geschäftsvorfälle eher zur Durchführung von detaillierten Prüfungen eignet, lassen sich Kennzahlenanalysen hier sinnvoll zur Beurteilung der Angemessen-

---

<sup>1</sup> Umschlagskennzahlen stellen Bestandsgrößen ins Verhältnis zu Erfolgsgrößen und geben an, wie häufig eine Vermögensposition in der betrachteten Periode umgeschlagen wurde (siehe ausführlich COENENBERG (2005), S. 989). Die wohl gängigsten Umschlagskennzahlen stellen die *Umschlagshäufigkeit der Vorräte* (siehe Kap. 2.2.2.3.4.2.2) und die *Umschlagshäufigkeit der Forderungen* (siehe Kap. 2.2.2.3.4.2.3) dar.

<sup>2</sup> Siehe hierzu auch ausführlich die Begriffsabgrenzung zu „analytischen Prüfungshandlungen“ in Kap. 2.1.1.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 76.

heit der Bewertung und der Klassifizierung anwenden.<sup>1</sup> Die wohl gebräuchlichste Kennzahl in diesem Prüffeld ist das Verhältnis zwischen dem Abschreibungsaufwand und den historischen Anschaffungskosten.<sup>2</sup> Sollte diese Kennzahl im Vergleich zum Vorjahr gesunken sein, könnte dies auf die Gefahr der unzureichenden Vornahme planmäßiger Abschreibungen hinweisen.<sup>3</sup> Bei anlageintensiven Industrieunternehmen könnte die Beweiskraft der Prüfungsnachweise dadurch erhöht werden, dass diese Kennzahl für homogene Gruppen mit vergleichbarer Nutzungsdauer und vergleichbaren Abschreibungsmethoden berechnet wird.<sup>4</sup>

Falls nachhaltige operative Probleme des geprüften Unternehmens dazu führen, dass bestimmte Vermögensgegenstände des Anlagevermögens nicht mehr eingesetzt werden, können Kennzahlenanalysen erste Anhaltspunkte auf eine potenzielle Unterauslastung geben, die bilanziell durch zusätzliche Abschreibungen der entsprechenden Maschinen o.Ä. zu berücksichtigen wären.<sup>5</sup> In diesem Zusammenhang könnte das Verhältnis zwischen fortgeführten Anschaffungs-/Herstellungskosten und den Umsatzerlösen ein geeignetes Mittel zur Beurteilung der Kapazitätsauslastung darstellen.<sup>6</sup> Sollte diese Kennzahl im Vergleich zum Vorjahr wesentlich gestiegen sein, könnte dies darauf hindeuten, dass weitere Abschreibungen auf bestimmte Vermögensgegenstände vorzunehmen sind.<sup>7</sup>

Weiterhin könnten Kennzahlenanalysen ein erhöhtes Fehlerrisiko bei der Klassifizierung von bestimmten Aufwandspositionen anzeigen. Sollte das Verhältnis zwischen den Instandhaltungsaufwendungen und den fortgeführten Anschaffungs-/Herstellungskosten im Zeitvergleich erheblich gesunken sein, könnte dies auf die fehlerhafte Aktivierung von Reparatur- und Erhaltungskosten zurückgeführt werden.<sup>8</sup>

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 28.

<sup>2</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 20.

<sup>3</sup> Als Nicht-Fehlerursache für eine derartige Entwicklung kommt allerdings auch in Betracht, dass das Sachanlagevermögen überwiegend aus älteren Vermögensgegenständen besteht, die degressiv abgeschrieben werden.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 81, 126-127.

<sup>5</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 127.

<sup>6</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 80.

<sup>7</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 80.

<sup>8</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 81.

### 2.2.2.3.4.2.2 Vorräte

Ein sinnvoller Einsatz von Kennzahlenanalysen im Bereich der Vorräte kann in Bezug auf die Beurteilung ihrer *Vollständigkeit, Bewertung, Klassifizierung und periodengerechten Abgrenzung* vorgenommen werden.<sup>1</sup> Eine erste Einschätzung der Vollständigkeit etwa, kann durch die Ermittlung der

$$\text{Vorratsintensität (\%)} = \frac{\text{Vorratsvermögen}}{\text{Gesamtvermögen}} \times 100$$

vorgenommen werden. Aufgrund des hohen Aggregationsgrads der in Beziehung gesetzten Jahresabschlussgrößen können hier jedoch nur erste Tendenzaussagen, etwa im Rahmen der Prüfungsplanung getroffen werden.<sup>2</sup> Eine höhere Aussagekraft könnte jedoch dadurch erreicht werden, dass beispielsweise bei einem Industrieunternehmen der jeweilige Anteil der Rohstoffe, unfertige oder fertige Erzeugnisse am Vorratsvermögen ermittelt und den entsprechenden Werten des Vorjahrs gegenübergestellt werden.<sup>3</sup>

Die wohl wichtigste Kennzahl zur Überprüfung der Vollständigkeit stellt hingegen die

$$\text{Umschlagshäufigkeit der Vorräte} = \frac{\text{Periodenabgang}}{\text{Ø Vorratsbestand}}$$

dar.<sup>4, 5</sup> Sie gibt an, wie häufig der durchschnittliche Vorratsbestand in der betrachteten Periode umgeschlagen wurde.<sup>6</sup> Sollte die Kennzahl gestiegen sein, könnte dies als ein erster Hinweis auf eine potenzielle Unterbewertung der Vorräte aufgrund unvollständiger Erfassung liegen.<sup>7</sup> Gründe hierfür könnten etwa darin liegen, dass während der Inventur bestimmte Bestände unabsichtlich nicht berücksichtigt wurden. Eine weitere Möglichkeit liegt darin, dass unverhältnismäßig hohe Inventurdifferenzen gebucht wur-

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 123-124.

<sup>2</sup> Vgl. MÜLLER (1996), S. 113.

<sup>3</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 20.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 77 i.V.m. COENENBERG (2005), S. 990. Eine alternative Kennzahl zur Umschlagshäufigkeit der Vorräte stellt die durchschnittliche Umschlagsdauer der Vorräte dar, die sich durch den Kehrwert der Umschlagshäufigkeit multipliziert mit 365 Tagen ergibt. Vgl. MCKEE (1989), S. 82.

<sup>5</sup> Bei der Verwendung des Umsatzkostenverfahrens entspricht der Periodenabgang den Umsatzkosten (cost of goods sold) und kann direkt der Gewinn- und Verlustrechnung entnommen werden. Basiert die Gewinn- und Verlustrechnung allerdings auf dem Gemeinkostenverfahren, kann der Periodenabgang nur indirekt ermittelt werden, etwa durch Berücksichtigung der Bestandsveränderungen oder durch die Hinzunahme von betrieblichen Aufzeichnungen. Als Surrogat für den Periodenabgang wird hier häufig auf die Umsatzerlöse zurückgegriffen. Für die Umschlagshäufigkeit der Vorräte siehe BAETGE/KIRSCH/THIELE (2004), S. 214 und für die Umschlagsdauer der Vorräte siehe MARTEN/QUICK/RUHNKE (2003), S. 280.

<sup>6</sup> Vgl. MCKEE (1989), S. 81.

<sup>7</sup> Vgl. MÜLLER (1996), S. 114-115.

den, um auf diese Weise einen geplanten Diebstahl in der abgelaufenen Periode zu vertuschen.<sup>1</sup> Allerdings könnte die Verringerung der Kennzahl auch auf Änderungen in der Geschäftstätigkeit des Unternehmens, z.B. dem Aufbau von Lagerbeständen aufgrund der Abkehr von einer rigiden „just-in-time“-Produktion, oder in dem wirtschaftlichen Umfeld, etwa der Verringerung der Einkaufspreise für Rohstoffe zurückzuführen sein.<sup>2</sup>

Die Umschlagshäufigkeit kann aber auch zur Einholung von Prüfungsnachweisen hinsichtlich der Bewertung der Vorräte herangezogen werden. So könnte eine unerwartete Verringerung dieser Kennzahl auf eine Überbewertung der Vorräte hindeuten. Gründe hierfür könnten etwa in der Nicht-Berücksichtigung von Abschlägen für überbevorratete, obsolete oder schlecht absetzbare Vorräte liegen.<sup>3</sup> Denkbar wäre auch die fehlende Durchführung von Abwertungen aufgrund niedrigerer Preise auf dem Beschaffungs- oder Absatzmarkt.<sup>4, 5</sup> Aber auch hier könnten Nicht-Fehlergründe für die Verringerung der Kennzahl verantwortlich sein, wie etwa ein allgemeiner Preisanstieg oder erhöhte Beschaffungsaktivitäten zum Jahresende aufgrund günstiger Konditionen.

Da die Umschlagshäufigkeit der Vorräte ähnlich wie zuvor der Forderungsumschlag auf sehr hoch verdichteten Jahresabschlussgrößen beruht, liefert sie allenfalls sehr schwache Prüfungsnachweise. Daher eignet sie sich eher für den Einsatz im Rahmen der Prüfungsplanung, etwa zur Feststellung von Prüfungsschwerpunkten, oder während der Prüfungsdurchführung als sog. preliminary analytical procedures.<sup>6</sup> Allerdings kann die Beweiskraft der Kennzahlenanalysen dadurch erhöht werden, dass detailliertere Größen gegenübergestellt werden. Zu denken wäre etwa an eine Ermittlung der

$$\text{Umschlagshäufigkeit der Rohstoffe} = \frac{\text{Periodenabgang}}{\text{Ø Rohstoffbestand}} \cdot 7$$

Analog könnte die Umschlagshäufigkeit auch auf unfertige und fertige Erzeugnisse bezogen werden.

Eine genaue Beurteilung der Angemessenheit der Bewertung könnte z.B. auch dadurch erreicht werden, dass man das Verhältnis der Abschläge für überbevorratete, obsolete

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 78.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 302-303.

<sup>3</sup> Vgl. MCKEE (1989), S. 82.

<sup>4</sup> Vgl. MÜLLER (1996), S. 159-160.

<sup>5</sup> Eine detaillierte Erörterung der Frage nach dem relevanten Markt für Gegenstände des Vorratsvermögens findet sich bei BAETGE/KIRSCH/THIELE (2005), S. 245-247 sowie QUICK/WOLZ (2007), S. 80.

<sup>6</sup> Siehe hierzu Kap. 2.1.2.2.1.2 bzw. 2.1.2.3.2.

<sup>7</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 77-78.

und schlecht absetzbare Vorräte zum gesamten Vorratsbestand mit dem Vorjahr vergleicht.<sup>1</sup> Eine wesentliche Verringerung könnte hier bereits einen starken Prüfungsnachweis für die Unterbewertung der Vorräte liefern und die Grundlage für die Korrekturbuchung bilden.

Eine relativ hohe Beweiskraft bringen Kennzahlenanalysen auch zur Beurteilung der Klassifizierung in Aufwendungen der laufenden Periode einerseits und aktivierungsfähigen Bestandteilen der Anschaffungs- und Herstellungskosten andererseits mit sich.<sup>2</sup> Eine fehlerhafte Klassifizierung im Bereich des Vorratsvermögens liegt beispielsweise vor, wenn Kosten für die Instandhaltung von Maschinen oder für Dienstreisen als Herstellungskosten der Vorräte aktiviert wurden.<sup>3</sup> Sinnvolle Kennzahlen für die Aufdeckung dieser Fehler sind die Verhältnisse der Material-, Fertigungs- und Gemeinkosten zum Vorratsbestand.<sup>4</sup>

#### 2.2.2.3.4.2.3 Forderungen

Der überwiegende Teil der Kennzahlenanalysen im Bereich der Forderungen zielt auf die Beurteilung ihrer *Einbringlichkeit*, *Vollständigkeit* sowie ihres *Vorhandenseins* und ihrer *periodengerechten Abgrenzung* ab.<sup>5</sup> So kann etwa eine gegenüber dem Vorjahr verringerte

$$\text{Forderungsintensität (\%)} = \frac{\text{Forderungsbestand}}{\text{Gesamtvermögen}} \times 100$$

einen ersten Hinweis darauf geben, ob bezüglich der Vollständigkeit der ausgewiesenen Forderungen mit einem erhöhten Fehlerrisiko zu rechnen ist.<sup>1</sup> Umgekehrt kann die Erhöhung dieser Kennzahl auf ein Risiko hinsichtlich des Vorhandenseins der ausgewiesenen Forderungen hindeuten. Aufgrund der möglichen Überbewertung der Forderungen birgt eine Erhöhung dieser Kennzahl im Zeitvergleich ein höheres Prüfungsrisiko als die Verringerung dieser Kennzahl.

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), 126.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 123.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 123.

<sup>4</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 20.

<sup>5</sup> Hierbei handelt es sich um eine im Jahresabschluss zum Ausdruck kommende Behauptung des Abschlusserstellers über die ökonomische Realität. Diese sog. Abschlusserklärungen sind vom Abschlussprüfer auf ihre Normenkonformität zu überprüfen. Eine ausführliche Übersicht über das System der durch ISA 300.17 geregelten Abschlusserklärungen geben SCHMIDT (2006), S. 4-6 und RUHNKE/LUBITZSCH (2006), S. 366-375.

Die wohl am häufigsten verwendete Kennzahl zur Beurteilung der Forderungen ist allerdings der

$$\text{Forderungsumschlag} = \frac{\text{Umsatzerlöse}}{\text{Forderungsbestand}} \cdot 2$$

Dabei weist der Forderungsumschlag darauf hin, wie häufig der durchschnittliche Forderungsbestand im Jahr zum Zufluss von liquiden Mitteln geführt hat.<sup>3</sup> Sollte sich diese Kennzahl im Zeitvergleich erhöht haben, könnte hierin ein erster Hinweis auf die Unvollständigkeit der ausgewiesenen Forderungen gesehen werden.<sup>4</sup> Dagegen könnte eine Verringerung dieser Kennzahl im Vergleich zum Vorjahr auf ein mögliches Problem in Bezug auf die Einbringlichkeit bzw. die Werthaltigkeit der Forderungen hindeuten.<sup>5</sup> Ein Grund für eine Senkung könnte etwa in der nicht (ausreichend) durchgeführten Wertberichtigung auf Forderungen liegen. Es könnten aber auch Sachverhalte zu dieser Entwicklung der Kennzahl geführt haben, die nicht auf Fehler im Jahresabschluss zurückzuführen sind, wie etwa die Verringerung der Anforderungen an die Bonität neuer Kunden, die Gewährung eines längeren Zahlungsziels, um neue Kunden zu erreichen oder bestehende Geschäftsbeziehungen zu pflegen, oder eine Krise in der entsprechenden Branche, die generell zu einer schlechteren Zahlungsmoral der Kunden führt.<sup>6</sup> Um den tatsächlichen Grund für die geschilderte Entwicklung festzustellen, sollte der Prüfer weitere Prüfungshandlungen vornehmen.

Da es sich bei dem Forderungsumschlag um eine auf sehr aggregierten Jahresabschlussgrößen basierende Kennzahl handelt, eignet sie sich insbesondere zum Einsatz in der Prüfungsplanung oder in der abschließenden Gesamtdurchsicht. Weniger sinnvoll erscheint ihre Anwendung hingegen zur Einholung von Prüfungsnachweisen im Rahmen der eigentlichen Prüfungsdurchführung. Hierfür sollte der Prüfer vielmehr auf Kennzahlen zurückgreifen, die auf detaillierten Jahresabschlussdaten beruhen. Beispielsweise könnte er das Verhältnis der Altersstrukturgruppen (etwa Forderungen, die seit 30, 60 und 90 Tagen nicht beglichen wurden) oder der wichtigsten Kunden zum gesamten

---

<sup>1</sup> Vgl. MÜLLER (1996), S. 182.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 302-303. Eine alternative Kennzahl zum Forderungsumschlag ist die durchschnittliche Umschlagsdauer von Forderungen, die sich durch den Kehrwert des Forderungsumschlags multipliziert mit 365 Tagen ergibt. Vgl. MCKEE (1989), S. 82.

<sup>3</sup> Vgl. MCKEE (1989), S. 82.

<sup>4</sup> Vgl. MÜLLER (1996), S. 186.

<sup>5</sup> Vgl. MCKEE (1989), S. 82.

<sup>6</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 302-303.

Forderungsbestand mit den entsprechenden Werten aus dem Vorjahr vergleichen.<sup>1</sup> Darüber hinaus könnte er auch die Summe der Forderungsabschreibungen bzw. der Wertberichtigungen auf Forderungen in Relation zu den Forderungen setzen. Eine Verringerung dieser Kennzahlen könnte ebenfalls auf ein Problem bezüglich der Werthaltigkeit oder der Einbringlichkeit hindeuten.<sup>2</sup>

#### **2.2.2.3.4.2.4 Verbindlichkeiten und Rückstellungen**

Zentrales Anliegen bei der Prüfung der Verbindlichkeiten ist die Beurteilung ihrer Vollständigkeit.<sup>3</sup> Hierzu ist der Einsatz von analytischen Prüfungshandlungen in Form von Kennzahlenanalysen besonders geeignet, da die Aufdeckung von nicht erfassten Verbindlichkeiten auch durch umfassende Belegprüfungen aufgrund ihres fehlenden Vorhandenseins nicht erreicht werden kann.<sup>4</sup> Ein erster noch sehr schwacher Nachweis für eine mögliche Unterbewertung der Verbindlichkeiten könnte durch die Ermittlung der

$$\text{Verbindlichkeitenintensität (\%)} = \frac{\text{Verbindlichkeiten}}{\text{Gesamtvermögen}} \times 100$$

bereitgestellt werden.<sup>5</sup> Sollte eine wesentliche Verringerung dieser Kennzahl zum Vorjahr identifiziert werden, könnte dies an der unvollständigen Verbuchung der Verbindlichkeiten liegen.<sup>6</sup> Allerdings ist die Aussagekraft dieser Prüfungshandlung sehr schwach und sollte daher eher im Rahmen der Prüfungsplanung zur Identifizierung von Prüfungsschwerpunkten eingesetzt werden.

Eine geeignetere Kennzahl zur Beurteilung der Vollständigkeit der Verbindlichkeiten ist in dem Verhältnis zwischen den Verbindlichkeiten und dem Gesamtaufwand der Periode (sog. Verbindlichkeitenumschlag) oder alternativ zwischen den Verbindlichkeiten

---

<sup>1</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 480.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 77; ARENS/ELDER/BEASLEY (2005), S. 480.

<sup>3</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 586-587.

<sup>4</sup> Siehe hierzu Kap. 2.1.2.3.1. Die einzige und wohl auch gebräuchlichste Einzelfallprüfung zur Identifizierung von nicht erfassten Verbindlichkeiten ist wohl die Inspektion der zu Beginn des neuen Geschäftsjahrs erfassten Verbindlichkeiten bzw. getätigten Auszahlungen. Bei dieser auch als „search for unrecorded account payables“ bezeichneten Vorgehensweise ist zu überprüfen, ob diese vom Mandanten dem Folgegeschäftsjahr zugeordneten Verbindlichkeiten nicht dem zu prüfenden Geschäftsjahr angehören. Vgl. ARENS/ELDER/BEASLEY (2005), S. 587 sowie BLOCHER/WILLINGHAM (1985), S. 127.

<sup>5</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 586.

<sup>6</sup> Vgl. MCKEE (1989), S. 84.

und dem Vorratsvermögen zu sehen.<sup>1</sup> Da den bei dieser Kennzahl einfließenden Jahresabschlussgrößen eine stärkere funktionale Beziehung als bei der Verbindlichkeitenintensität zu attestieren ist, weisen hierauf basierende Erkenntnisse tendenziell eine höhere Beweiskraft auf. Allerdings ist auch hier anzumerken, dass diese Kennzahl auf relativ hoch verdichteten Größen basiert, was ihre Aussagekraft schwächt.

Um Kennzahlen zu ermitteln, die auf weniger verdichteten Jahresabschlussgrößen basieren, könnte man die Verbindlichkeiten auch nach bestimmten Merkmalen, etwa nach der Alterstruktur, unterteilen und in Verhältnis zu den Gesamtverbindlichkeiten setzen.<sup>2</sup> Anstelle der gesamten Verbindlichkeiten könnte man auch auf den Gesamtaufwand der Periode oder auf den Vorratsbestand zurückgreifen. Weitere geeignete Merkmale zur Unterteilung der Verbindlichkeiten stellen auch Größenklassen oder die Zugehörigkeit der Lieferanten zu bestimmten ABC-Klassen (gemessen am Gesamteinkauf in der Periode) dar.

Auch im Bereich der Rückstellungen können Kennzahlenanalysen starke und kostengünstige Prüfungsnachweise zur Verfügung stellen. Ihr Einsatz ist besonders dort sinnvoll, wo die Bildung von Rückstellungen auf gesetzliche Bestimmungen zurückzuführen ist.<sup>3</sup> Beispielsweise kann ein Rückgang des Verhältnisses zwischen den Rückstellungen für Garantie- oder Produkthaftung und den Umsatzerlösen eine Unterbewertung indizieren.<sup>4</sup> Weitere nützliche Kennzahlen aufgrund vergleichbarer Ursache-Wirkungs-Zusammenhänge könnten die

- Prozesskosten-Rückstellungen und Umsatzerlöse,
- Rückstellungen für Überstunden sowie nicht in Anspruch genommene Urlaubsansprüche und die Lohnkosten der Periode bzw. der Mitarbeiterzahl,
- Rückstellungen für Schadenersatzansprüche und Umsatzerlöse oder
- Rückstellungen für Instandhaltung und Instandhaltungsaufwendungen in der Periode bzw. historische Anschaffungs-/Herstellungskosten des Sachanlagevermögens

in Verhältnis zueinander gesetzt werden.

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 127 sowie FITZSIMONS/LEVINE (1989), S. 20.

<sup>2</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 20.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 127.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 127.



### 2.2.2.3.4.3 Gewinn- und Verlustrechnung

#### 2.2.2.3.4.3.1 Beurteilung der Umsatzerlöse

Ein sinnvoller Einsatz von analytischen Prüfungshandlungen im Rahmen der Gewinn- und Verlustrechnung kann sowohl auf der Ertrags- als auch auf der Aufwandsseite erfolgen.<sup>1</sup> Der zentrale Posten zur Beurteilung der ausgewiesenen Erträge ist in den Umsatzerlösen zu sehen. Die Beurteilung der Umsatzerlöse ist auf zwei grundlegende Aspekte ausgerichtet.<sup>2</sup> Zum einen hat der Prüfer das *Vorhandensein* der ausgewiesenen Umsatzerlöse zu prüfen. Hierbei sollte er sowohl eine fehlerhafte Periodenabgrenzung als auch einen Ausweis fingierter Verkäufe in Betracht ziehen. Zum anderen obliegt es dem Prüfer, die *Vollständigkeit* der ausgewiesenen Umsatzerlöse zu prüfen. Eine durch nicht erfasste Verkäufe erwünschte Gewinnminderung könnte etwa angestrebt werden, um weniger Steuern abführen zu müssen oder um die Gewinne im Zeitablauf zu glätten.

Erste Anhaltspunkte zur Normenkonformität der Umsatzerlöse können sich bereits im Rahmen der Prüfung anderer Prüffelder ergeben. Hierbei handelt es sich insbesondere um den Bereich der Vorräte und der Forderungen.<sup>3</sup> Eine ungewöhnliche Verringerung des *Forderungsumschlags* könnte etwa ein Indiz für eine fehlerhafte Periodenabgrenzung von Umsätzen oder fiktive Verkaufstransaktionen sein.<sup>4, 5</sup> Eine unerwartete Verringerung (Erhöhung) der *Umschlagshäufigkeit der Vorräte* könnte hingegen auf eine fehlende (doppelte) Erfassung von Periodenabgängen zurückzuführen sein.<sup>6</sup> Daher erscheint es durchaus geeignet, die Beurteilung der Umsatzerlöse gleichzeitig mit der Forderungs- oder auch Vorratsprüfung vorzunehmen.<sup>7</sup>

Eine sehr weit verbreitete Kennzahl zur Überprüfung der Umsatzerlöse stellt die

---

<sup>1</sup> Vgl. GILL/COSSERAT/LEUNG/CORAM (2001), S. 353; ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>2</sup> Vgl. im Folgenden BLOCHER/WILLINGHAM (1985), S. 130.

<sup>3</sup> Vgl. im Folgenden BLOCHER/WILLINGHAM (1985), S. 130. Zum Einsatz von Kennzahlenanalysen im Bereich der Forderungen und Vorräte siehe ausführlich Kap. 2.2.2.3.4.2 und 2.2.2.3.4.2.2.

<sup>4</sup> Hierbei wird unterstellt, dass der Zähler „Umsatzerlöse“ betragsmäßig höher ausfällt als der Nenner „Forderungen“, da nur für diesen Fall gilt, dass eine Erhöhung der Umsatzerlöse und Forderungen um den identischen Betrag zu einer proportional stärkeren Erhöhung des Nenners und somit zu einer Verringerung der Kennzahl „Umschlagshäufigkeit der Forderungen“ führt.

<sup>5</sup> Allerdings ist hier darauf hinzuweisen, dass fingierte Umsätze nicht zwangsläufig zu einer Verringerung der Kennzahl führen. Vielmehr können fingierte Umsätze auch zu einer Erhöhung dieser Kennzahl führen, falls fingierte Barverkäufe vorliegen, ein fingierter Forderungsausgleich vor dem Bilanzstichtag stattgefunden hat oder eine fingierte Abschreibung der entsprechenden Forderungen durchgeführt wurde.

<sup>6</sup> Ein nicht erfasster (doppelt erfasster) Periodenabgang führt zu einer Unterbewertung (Überbewertung) des Zählers „Periodenabgang“ und einer Überbewertung (Unterbewertung) des Nenners „durchschnittlicher Vorratsbestand“, was zwingend eine Verringerung (Erhöhung) der Kennzahl „Umschlagshäufigkeit der Vorräte“ bewirkt.

<sup>7</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 130.

$$\text{Bruttogewinnmarge (\%)} = \frac{\text{Bruttoergebnis vom Umsatz (Rohgewinn)}}{\text{Umsatzerlöse}} \times 100$$

dar.<sup>1</sup> Eine unerwartete Erhöhung (Verringerung) dieser Kennzahl könnte auf einer Überbewertung (Unterbewertung) der Umsätze basieren.<sup>2</sup> Somit dient diese Kennzahl sowohl der Beurteilung der Vollständigkeit als auch des Vorhandenseins der Umsätze. Einen höheren Detaillierungsgrad und somit stärkere Aussagekraft könnte dadurch erzielt werden, dass man die Bruttogewinnmarge für einzelne Produktgruppen ermittelt und mit denen des Vorjahrs vergleicht.<sup>3</sup>

Eine weitere auf den gleichen Zweck gerichtete Kennzahl ist das Verhältnis zwischen den Warenrücknahmen sowie Gutschriften zu den Umsatzerlösen. Auch hier könnten wesentliche Abweichungen auf einen Fehler in der Umsatzerfassung hindeuten.<sup>4</sup> So könnte eine unerwartete Verringerung der Kennzahl auf eine fehlerhafte Periodengrenzung der Umsätze oder auf fingierte Verkäufe zurückzuführen sein. Eine unerwartete Erhöhung hingegen könnte in der fehlenden Erfassung von Verkäufen begründet liegen.

#### **2.2.2.3.4.3.2 Beurteilung der Erträge aus Zinsen und Dividenden**

Bestimmte Unternehmen erwirtschaften ihre Gewinne nicht nur aus Warenverkäufen, sondern auch aus erhaltenen Zinsen und Dividenden. In diesen Fällen können die erzielten Erträge ebenfalls sinnvoll auf ihre Plausibilität untersucht werden, indem sie ins Verhältnis zu denjenigen Jahresabschlussgrößen gesetzt werden, die die Grundlage für die Erträge bilden. Dabei können einerseits die Zinserträge den Ausleihungen und die Dividendenerträge den Beteiligungen gegenübergestellt werden.<sup>5</sup> Sollten hier ebenfalls unerwartete Abweichungen festgestellt werden, könnte dies vom Prüfer zum Anlass genommen werden, von einer erhöhten Fehlerwahrscheinlichkeit in den entsprechenden Erträgen auszugehen.

---

<sup>1</sup> Vgl. etwa ARENS/ELDER/BEASLEY (2005), S. 480; MCKEE (1989), S. 90.

<sup>2</sup> Da die Umsatzerlöse immer höher ausfallen als der Rohgewinn, steigt (sinkt) der Rohgewinn im Zähler bei einer betragsgleichen Erhöhung (Verringerung) beider Positionen proportional stärker als die Umsatzerlöse im Nenner. Dieses führt zwingend zu einer Erhöhung (Verringerung) der Kennzahl „Bruttogewinnmarge“.

<sup>3</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 21; BLOCHER/WILLINGHAM (1985), S. 130.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 131; ARENS/ELDER/BEASLEY (2005), S. 480.

<sup>5</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 80-81.

### 2.2.2.3.4.3.3 Beurteilung der Aufwendungen

Zur Überprüfung der Aufwendungen kommt im Gegensatz zur Überprüfung der Erträge eine Vielzahl von Aufwandskonten in Betracht. Das zentrale Anliegen bei der Untersuchung der einzelnen Aufwandspositionen liegt in der Beurteilung der Vollständigkeit. Hierzu stehen dem Prüfer die drei folgenden Arten von Kennzahlenanalysen zur Verfügung. Die erste Art zeigt die jeweils zu betrachtende Aufwandsposition als einen Prozentsatz von den gesamten Umsatzerlösen an.<sup>1</sup> Dieser im Folgenden als Strukturanalyse<sup>2</sup> bezeichneten Vorgehensweise ist ein hoher Stellenwert beizumessen, weil sie die zugrunde liegenden wirtschaftlichen Zusammenhänge reflektiert.<sup>3</sup> So kann es beispielsweise sinnvoll sein

- die Warenrückläufe und Gutschriften,
- die Kosten der umgesetzten Einheiten,
- die Vertriebskosten,
- die Forderungsabschreibungen,
- die Kosten für Garantiefälle oder
- die Verbrauchssteuern

in Relation zu den Umsätzen zu betrachten.<sup>4</sup>

Derartige Strukturanalysen sind im Bereich der Gewinn- und Verlustrechnung deshalb relevant, weil der überwiegende Teil der Aufwandspositionen in einer stabilen Relation zu den erzielten Umsätzen steht. Denn bei einer Steigerung der Umsatzerlöse werden sich die meisten Aufwandskonten in einer abschätzbaren Art und Weise ebenfalls erhöhen.<sup>5</sup> Daher ist es durchaus sinnvoll, derartige Kennzahlen auf wesentliche Änderungen gegenüber dem Vorjahr zu untersuchen, um auf diesem Wege Konten mit erhöhtem Fehlriskio zu identifizieren.

---

<sup>1</sup> Vgl. MCKEE (1989), S. 90.

<sup>2</sup> im angloamerikanischen Raum auch als *common size ratios* oder *common size analysis* genannt. Vgl. etwa MCKEE (1989), S. 91; KNECHEL (2001), S. 274.

<sup>3</sup> Vgl. MCKEE (1989), S. 91.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 82; FITZSIMONS/LEVINE (1989), S. 21; MCKEE (1989), S. 90.

<sup>5</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 82. Strukturanalysen lassen sich zwar auch zur Beurteilung von einzelnen Bilanzpositionen heranziehen, indem etwa die einzelnen Konten des Gesamtvermögens oder des Gesamtkapitals als Prozentsatz des Gesamten dargestellt werden (vgl. etwa KNECHEL (2001), S. 274; MCKEE (1989), S. 91). Allerdings sind die Zusammenhänge hier nicht so aussagekräftig, weil

Allerdings sollte sich der Prüfer in diesem Zusammenhang darüber bewusst sein, dass sich einzelne Aufwandskonten bezüglich ihres Anteils fixer und variabler Bestandteile voneinander unterscheiden können.<sup>1</sup> Sollten die betrachteten Aufwendungen überwiegend variablen Charakter aufweisen, ist deren Entwicklung bei gleichzeitiger Änderung der Umsätze sicherlich einfacher zu bestimmen als bei Aufwendungen, die überwiegend auf fixen Bestandteilen beruhen. Daher sollte der Prüfer bei Strukturanalysen in der Gewinn- und Verlustrechnung zunächst einen ausreichenden Kenntnisstand über das Verhältnis der fixen zu den variablen Bestandteilen der zu beurteilenden Aufwendungen erlangen.<sup>2</sup>

Bei der zweiten Art von Kennzahlenanalysen im Bereich der Aufwendungen werden Aufwandspositionen und Bilanzkonten gegenübergestellt, die einen engen Bezug zueinander haben und daher ein stabiles Verhältnis aufweisen.<sup>3</sup> Beispielsweise erscheint es durchaus sinnvoll,

- die Zinsaufwendungen und die Verbindlichkeiten gegenüber Kreditinstituten,
- die Abschreibungsaufwendungen und das Sachanlagevermögen,
- die Instandhaltungsaufwendungen und das Sachanlagevermögen,
- die Versicherungsaufwendungen und die Vorräte zuzüglich des Sachanlagevermögens,
- die Forderungsabschreibungen und den Forderungsbestand

gegenüber zu stellen und auf auffällige Abweichungen zum Vorjahr zu inspizieren.<sup>4</sup>

Die dritte Art von Kennzahlenanalysen im Aufwandsbereich setzt einzelne Aufwendungen in Relation zu anderen Konten der Gewinn- und Verlustrechnung, die einen engen Bezug zu den zu prüfenden Aufwendungen aufweisen. Zu denken wäre hier etwa an eine Gegenüberstellung

- der Sozialversicherungskosten und der Lohnkosten,

---

Bilanzpositionen u.a. auch nicht operative Elemente beinhalten. Vgl. BLOCHER/WILLINGHAM (1985), S. 82.

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 131.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 83.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 80.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 131-132; FITZSIMONS/LEVINE (1989), S. 20-21.

- der Steuern vom Einkommen und Ertrag und des Jahresüberschusses vor Steuern und
- der Material-, Fertigungseinzel- oder -gemeinkosten und den gesamten Herstellungskosten.<sup>1</sup>

#### 2.2.2.3.5 Probleme/Grenzen der Kennzahlenanalyse

Obwohl dem Prüfer mit Kennzahlenanalysen ein nützliches Verfahren zur Beurteilung der Plausibilität von Jahresabschlussdaten zur Verfügung steht, darf er ihre Effizienz und Effektivität nicht überschätzen. Vielmehr sollte sich der Prüfer auch über die konzeptionellen Schwächen von Kennzahlenanalysen bewusst sein, denn nur so ist er in der Lage, die Aussagekraft seiner erlangten Prüfungsnachweise in den einzelnen Prüfungssituationen richtig einzuschätzen und seine weitere Vorgehensweise angemessen fortzuführen.

Das größte Problem von Kennzahlenanalysen ist in ihren Optimalitätsbedingungen zu sehen. Demnach ist ihre Aussagekraft immer dann als besonders hoch einzuschätzen, wenn ein möglichst

- *streng linearer* und
- *monokausaler*

Zusammenhang zwischen den ins Verhältnis zueinander gesetzten Jahresabschlussgrößen besteht.<sup>2</sup> Diese beiden Bedingungen werden jedoch in der Realität faktisch so gut wie nie erfüllt.<sup>3</sup>

Nimmt man beispielsweise an, dass sich die Vertriebsaufwendungen sowohl aus variablen als auch aus fixen Bestandteilen zusammensetzen, so müsste die lineare Gleichung wie folgt formuliert werden:<sup>4</sup>

---

<sup>1</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 81, 131-132; FITZSIMONS/LEVINE (1989), S. 20-21; ARENS/ELDER/BEASLEY (2005), S. 211.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 301.

<sup>3</sup> Vgl. LEFFSON (1984), S. 171.

<sup>4</sup> Vgl. im Folgenden BLOCHER/WILLINGHAM (1985), S. 85.

$$(1) \quad y = a + kx$$

mit  $y$  = Vertriebsaufwendungen

$a$  = fixer Bestandteil der Vertriebsaufwendungen

$k$  = Kennzahl

$x$  = Umsatzerlöse

Löst man diese Gleichung nach  $k$  auf, würde sich folgende Gleichung ergeben:

$$(2) \quad k = y/x - a/x .$$

Für eine streng lineare Beziehung zwischen den im Rahmen von Kennzahlenanalysen zueinander in Beziehung gesetzten Jahresabschlussgrößen (für das vorliegende Beispiel sind es die Vertriebsaufwendungen und die Umsatzerlöse) müsste der fixe Bestandteil (a) gleich null sein, um

$$(3) \quad k = y/x$$

zu erhalten und die Optimalitätsbedingung zu erfüllen. Sind jedoch fixe Bestandteile enthalten, stellt der zweite Quotient in der Gleichung (2), also  $(-a/x)$ , den Störfaktor dar und die Kennzahl  $k$  wäre für unterschiedliche Ausprägungen der Umsatzerlöse nicht mehr vergleichbar.

Hinzu kommt, dass Jahresabschlussgrößen in der Regel nicht ausschließlich von einer einzigen anderen Jahresabschlussgröße beeinflusst werden. Somit kann kaum von einem monokausalen Zusammenhang ausgegangen werden.<sup>1</sup> Die Höhe des Forderungsbestands etwa wird wesentlich von den Umsatzerlösen beeinflusst. Dieser Zusammenhang kommt in der Kennzahl des Forderungsumschlags zum Ausdruck.<sup>2</sup> Nimmt diese Kennzahl im Vergleich zum Vorjahr unerwartet stark ab, könnte dies auf eine Überbewertung der Forderungen zurückzuführen sein. Es ist aber auch denkbar, dass das Unternehmen seinen Kunden ein längeres Zahlungsziel gewährt hat, dass sich die Zahlungsmoral der Kunden aufgrund wirtschaftlicher Probleme verschlechtert hat oder dass gerade zum Jahresende besonders große Umsätze getätigt wurden.<sup>3</sup>

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 133-134.

<sup>2</sup> Vgl. hierzu Kap. 2.2.2.3.4.2.

<sup>3</sup> Vgl. LEFFSON (1984), S. 168.

Ähnlich kann es sich mit den Vertriebskosten verhalten.<sup>1</sup> Diese werden maßgeblich von den Umsatzerlösen beeinflusst. Allerdings können sich darüber hinaus auch das Verkaufssortiment und die damit entstehenden Transport- und Verpackungskosten auf die Höhe der Vertriebskosten auswirken.

Des Weiteren kann die Vergleichbarkeit von Kennzahlen auch erheblich durch Änderungen in den vom Mandanten herangezogenen Bewertungsgrundsätzen beeinträchtigt werden.<sup>2, 3</sup> Derartige Unterschiede könnten etwa im Rahmen der Vorräte hinsichtlich der Verbrauchsfolgebewertung oder der Einbeziehung von Gemeinkosten in die Herstellungskosten vorliegen oder aufgrund von Ermessensspielräumen, etwa bei Schätzungen in Bezug auf die Wertberichtigung von Forderungen oder die Abschreibungen von überbevorrateten, obsoleten bzw. unverkäuflichen Gegenständen des Vorratsvermögens.<sup>4</sup>

Schließlich ist eine potenzielle Gefahr von Kennzahlenanalysen auch darin zu sehen, dass sich die Auswirkung einer Abweichung im Zähler und die Auswirkung einer Abweichung im Nenner in Bezug auf den Quotienten kompensieren.<sup>5</sup> Somit besteht generell die Gefahr, dass beide Jahresabschlussgrößen wesentliche Fehler enthalten, der Prüfer aber keine erhöhte Fehlerwahrscheinlichkeit erkennt, weil die Kennzahl keine Abweichung signalisiert.<sup>6</sup> In diesen Fällen wäre die Effektivität der Jahresabschlussprüfung nachhaltig beeinträchtigt.

Die aufgeführten Schwächen der Kennzahlenanalyse sollen jedoch nicht zum Anlass genommen werden, den Prüfer vom Einsatz analytischer Prüfungshandlungen in Form von Kennzahlenanalysen abzuhalten. Der Prüfer sollte vielmehr versuchen, die Kennzahlenanalyse nur auf solche Jahresabschlussgrößen anzuwenden, die in einem sachlogischen Zusammenhang zueinander stehen und nur geringfügig von weiteren Faktoren beeinflusst werden. Berücksichtigt er dabei überwiegend stetige Jahresabschlussgrößen, die stets in gleicher Weise in die Kennzahl eingeflossen sind, und bringt er bei der In-

---

<sup>1</sup> Vgl. im Folgenden BLOCHER/WILLINGHAM (1985), S. 87.

<sup>2</sup> Vgl. WILSON/COLBER (1991), S. 40; GUY/WINTERS/ALDERMAN (1999), S. 378.

<sup>3</sup> Allerdings sind Änderungen in den Bewertungsgrundsätzen durch den Grundsatz der Bewertungsstetigkeit (§ 252 Abs. 1 Nr. 6 HGB, IAS 8.15) nur unter bestimmten, z.T. sehr restriktiven Voraussetzungen zulässig (§ 252 Abs. 2 HGB, IAS 8.14).

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 84.

<sup>5</sup> Vgl. MÜLLER (1996), S. 44.

<sup>6</sup> In Bezug auf die Bruttogewinnmarge etwa könnte eine Überbewertung (Unterbewertung) der Umsatzerlöse durch eine Überbewertung (Unterbewertung) der Umsatzkosten kompensiert werden.

terpretation der Ergebnisse zudem die nötige Sorgfalt auf, so ist in der Kennzahlenanalyse ein durchaus wirksames und wirtschaftliches Verfahren zu sehen.<sup>1</sup>

### **2.2.3 Informationen aus dem aktuellen Geschäftsjahr**

#### **2.2.3.1 Benchmark-Vergleiche**

##### **2.2.3.1.1 Übersicht über die Arten von Benchmark-Vergleichen**

Bei den Benchmark-Vergleichen stellt der Prüfer den zu beurteilenden Jahresabschlussdaten einen Wert (sog. *Benchmark*) gegenüber, von dem er erwartet, dass er einen geeigneten Richtwert für den tatsächlichen Wert der untersuchten Jahresabschlussposition darstellt. Dabei kann generell zwischen den folgenden Arten von Benchmark-Vergleichen differenziert werden:<sup>2</sup>

- *Planzahlen-Benchmarking*;
- *Branchen-Benchmarking*;
- *Betriebs-Benchmarking*.

##### **2.2.3.1.2 Planzahlen-Benchmarking**

Eine Vielzahl von Unternehmen verwendet Planzahlen zur verbesserten Steuerung und Kontrolle ihrer Geschäftstätigkeit.<sup>3</sup> Da es sich bei Planzahlen um Erwartungen des Mandanten hinsichtlich seiner operativen und finanziellen Ergebnisse handelt, sollte der Prüfer auch diese Informationsquelle als Vergleichsmaßstab zur Beurteilung der Plausibilität der ausgewiesenen Jahresabschlussdaten heranziehen.<sup>4</sup> Sollte er dabei auf wesentliche Abweichungen zwischen den Planzahlen des Unternehmens einerseits und den ausgewiesenen Jahresabschlusswerten andererseits treffen, könnte dies auf eine erhöhte Fehlerwahrscheinlichkeit hindeuten.<sup>1</sup> Vergleiche zwischen Planzahlen und Jahresabschlusswerten sind insbesondere bei der Prüfung von öffentlichen Unternehmen relevant, da Planzahlen bzw. Budgetierungen von öffentlichen Mitteln aufgrund gesetzli-

---

<sup>1</sup> So auch MÜLLER (1996), S. 44.

<sup>2</sup> Vgl. WALLACE (1983a), S. 31.

<sup>3</sup> Vgl. JOHNSON/JOHNSON (1997), S. 83; MARTEN/QUICK/RUHNKE (2007), S. 309; ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>4</sup> Vgl. KNECHEL (2001), S. 281; MCKEE (1982), S. 39.



cher Vorgaben oder behördlicher Beschlüsse eine herausragende Rolle einnehmen.<sup>2</sup> Aber auch bei Prüfungen von privaten Unternehmen können analytische Prüfungshandlungen anhand von Planzahlen eingesetzt werden, um etwa Maßnahmen zur Gewinnglättung im letzten Geschäftsquartal aufzudecken.<sup>3</sup>

Ein wesentlicher Vorteil bei der Verwendung von Planzahlen besteht darin, dass sie bereits ungewöhnliche, jedoch antizipierte Transaktionen im Geschäftsjahr umfassen.<sup>4</sup> Sollte das Unternehmen beispielsweise die Schließung eines Unternehmensbereichs, den Rückzug aus einem bestimmten Marktsegment oder die Änderung eines Produktionssortiments beabsichtigen und dieses bereits in seinen Planzahlen berücksichtigt haben, so stellen Planzahlen unter der Bedingung einer angemessenen und ordentlichen Unternehmensplanung sicherlich einen verlässlicheren Erwartungswert als etwa der Vorjahreswert dar.<sup>5</sup>

Ein weiterer Vorteil der Verwendung von Planzahlen liegt darin, dass viele Unternehmen die Abweichungen von den Planzahlen bereits im Rahmen ihrer Kontrollprozeduren analysiert und ausreichend erläutert haben.<sup>6</sup> Der Prüfer sollte sich derartige Unterlagen einholen und beurteilen, ob die Erklärungen angemessen erscheinen.<sup>7</sup>

Kritisch sind Planzahlen allerdings dann zu betrachten, wenn sie nicht mit der angemessenen Sorgfalt erstellt wurden und unrealistisch erscheinen. Der Prüfer sollte sich daher zunächst mit dem Erstellungsprozess der Planzahlen vertraut machen und dabei insbesondere beurteilen, ob die hierfür verantwortlichen Personen über die ausreichende Erfahrung und Kompetenz im Umgang mit Planzahlen verfügen.<sup>8</sup> Weiterhin sollte der Prüfer ausschließen können, dass die Planzahlen unrealistische Zielvorgaben darstellen, welche lediglich zum Zwecke der Motivation von Mitarbeitern eingesetzt wurden und somit nicht in Einklang mit den Zielen des Jahresabschlusses stehen.<sup>9</sup> In derartigen Fällen erbringen Planzahlen lediglich schwache Prüfungsnachweise.<sup>10</sup>

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 309; ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>2</sup> Vgl. JOHNSON/JOHNSON (1997), S. 39 m.w.N.

<sup>3</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 617; BLOCHER/PATTERSON (1996), S. 54; SCOTT/WALLACE (1994), S. 64.

<sup>4</sup> Vgl. GAUNTT/GLEZEN (1997), S. 56.

<sup>5</sup> Vgl. GAUNTT/GLEZEN (1997), S. 56-57.

<sup>6</sup> Vgl. MCKEE (1982), S. 39.

<sup>7</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 8-9.

<sup>8</sup> Vgl. KAYADELEN (2006), S. 572; ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>9</sup> Vgl. MÜLLER (1996), S. 54 m.w.N.

<sup>10</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 309.

Ein weiteres Problem des Planzahlen-Benchmarkings besteht darin, dass der Mandant die Jahresabschlusswerte möglicherweise dahingehend manipuliert, dass sie weitestgehend mit den Planzahlen übereinstimmen.<sup>1</sup> Folglich würde ein Vergleich trotz fehlerhafter Jahresabschlusssdaten dazu führen, dass der Prüfer keine wesentliche Abweichung entdecken und die Manipulation nicht feststellen kann.<sup>2</sup> Um dieses Risiko abschätzen zu können, sollte sich der Prüfer darüber informieren, ob ausreichende Kontrollen zur Vermeidung von manipulativen Eingriffen durch das Management existieren und effektiv sind, d.h. nicht oder kaum von dem Management umgangen werden können.<sup>3</sup> Dieses sollte regelmäßig Bestandteil der Prüfung des internen Kontrollsystems sein.<sup>4</sup>

### **2.2.3.1.3 Branchen-Benchmarking**

#### **2.2.3.1.3.1 Wesen des Branchen-Benchmarkings**

Beim Branchen-Benchmarking greift der Prüfer zur Entwicklung des Erwartungswerts nicht wie bei den bisher vorgestellten Verfahren auf Informationen des gleichen Unternehmens zurück, sondern auf Informationen von anderen Unternehmen der gleichen Branche.<sup>5</sup> Als Branche kann dabei eine Gruppe von Unternehmen bezeichnet werden, die ähnliche Produkte auf dem gleichen Markt anbieten, vergleichbare Rohstoffe und Produktionsverfahren verwenden und darüber hinaus über ähnliche Distributionswege verfügen.<sup>6</sup> Da Unternehmen der gleichen Branche eher als Unternehmen unterschiedlicher Branchen miteinander verglichen werden können,<sup>7</sup> werden die Branchenwerte zur Bildung einer normativen Größe herangezogen, die als Richt- bzw. Erwartungswert für das zu prüfende Unternehmen dient.<sup>8</sup> Anschließend werden die ungeprüften Buchwerte dem ermittelten Richt- bzw. Erwartungswert gegenübergestellt und auf Abweichungen untersucht.<sup>9</sup> Sollte der Prüfer dabei eine wesentliche unerwartete Abweichung feststel-

---

<sup>1</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>2</sup> Vgl. MÜLLER (1996), S. 54; MÜLLER/KROPP (1992), S. 154.

<sup>3</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 213.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 279.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S.308. Häufig spricht man in diesem Zusammenhang auch von überbetrieblichen Vergleichen. Vgl. etwa MÜLLER (1996), S. 53-53.

<sup>6</sup> Vgl. MCKEE (1989), S. 102.

<sup>7</sup> Vgl. MCKEE (1989), S. 102-103.

<sup>8</sup> Vgl. MCKEE (1989), S. 78, 102.

<sup>9</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 308; MCKEE (1989), S. 78.

len, könnte er dies zum Anlass nehmen, von einem erhöhten Fehlerrisiko im Datenmaterial des geprüften Unternehmens auszugehen.<sup>1</sup>

Die Vergleichswerte beim Branchen-Benchmarking können einerseits auf einzelnen Jahresabschlusspositionen, wie etwa den Vorräten, Forderungen oder Rückstellungen, basieren, und andererseits auf jahresabschlussbezogenen Kennzahlen. Dabei kann der Prüfer auf alle bereits im Rahmen der zeitbezogenen Kennzahlenanalyse (sog. *time series approach*) vorgestellten Kennzahlen zurückgreifen<sup>2</sup>, da davon ausgegangen wird, dass einzelne Ursache-Wirkungszusammenhänge innerhalb der Jahresabschlussdaten nicht nur über mehrere Perioden beim gleichen Unternehmen, sondern auch bei mehreren, der gleichen Branche angehörenden Unternehmen stabil sind.<sup>3</sup> Positiv hervorzuheben sind Kennzahlen auch deshalb, weil sie bei Unternehmen unterschiedlicher Größe robuster sind als absolute Zahlen.<sup>4</sup>

#### **2.2.3.1.3.2 Arten des Branchen-Benchmarkings**

##### **2.2.3.1.3.2.1 Branchendurchschnittswert-Benchmarking**

Dem Prüfer stehen beim Branchen-Benchmarking zwei unterschiedliche Arten von Vergleichsmöglichkeiten zur Verfügung. Dabei handelt es sich einerseits um das *Branchendurchschnittswert-Benchmarking* und andererseits um das *Unternehmens-Benchmarking*.<sup>5</sup>

Bei der ersten Art des Branchen-Benchmarkings verwendet der Prüfer Branchendurchschnittswerte, die aus Daten mehrerer Unternehmen ermittelt werden.<sup>6</sup> Zur Erlangung der benötigten Branchendurchschnittswerte kann der Prüfer auf unterschiedliche Quellen zurückgreifen. In Betracht kommen etwa Berichte von statistischen Ämtern, Banken und Finanzanalysten, Mitteilungen von Industrie- und Handelskammern, Unternehmensverbänden und Wirtschaftsvereinigungen oder die Konsultation von Branchenspezialisten innerhalb der eigenen Prüfungsgesellschaft.<sup>7</sup> Falls der Zugang zu einigen Quel-

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 308.

<sup>2</sup> Vgl. hierzu ausführlich Kap. 2.2.2.3.

<sup>3</sup> Vgl. MCKEE (1989), S. 102; BLOCHER/WILLINGHAM (1985), S. 71-72.

<sup>4</sup> Vgl. MCKEE (1989), S. 91; BLOCHER/WILLINGHAM (1985), S. 72.

<sup>5</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 308; ARENS/ELDER/BEASLEY (2005), S. 210.

<sup>6</sup> Neben dem Durchschnitt eignen sich auch andere Parameter, wie etwa der Median, zur Verdichtung von Daten mehrerer Unternehmen. Einen Überblick hierzu gibt BLOCHER/WILLINGHAM (1985), S. 90-91.

<sup>7</sup> Vgl. IDW (2006), R 383.

len u.U. ausschließlich Mitgliedern vorbehalten sein sollte, könnte der Prüfer versuchen, den Mandanten in die Datenbeschaffung einzubeziehen.<sup>1</sup> Dies sollte der Prüfer insbesondere bei Branchen beachten, die einer dynamischen Umwelt ausgesetzt sind.<sup>2</sup>

Obwohl Branchendurchschnittswerte ein nützliches Vergleichsobjekt bei der Durchführung von analytischen Prüfungshandlungen sind, sollte sich der Prüfer auch über die Probleme ihrer Anwendung bewusst sein. So kann es dem Prüfer bereits erste Schwierigkeiten bereiten, das zu prüfende Unternehmen einer bestimmten Branche zuzuordnen. Dies mag zwar weniger der Fall sein, wenn es sich dabei um ein Unternehmen handelt, welches ausschließlich eine bestimmte Art von Produkten oder Dienstleistungen anbietet. Häufig liegen auch Unternehmenskonglomerate vor, die ein breit gefächertes Produkt- und Serviceangebot aufweisen,<sup>3</sup> und somit nicht in eine bestimmte Branche hineinpassen.<sup>4</sup> Ebenso problematisch gestaltet sich die Zuordnung, wenn das Unternehmen über ein einzigartiges, sich in wesentlichen Eigenschaften von den Produkten der übrigen Branche differenzierendes Angebot verfügt.<sup>5</sup>

Ein zweites wesentliches Problem liegt darin, dass sich Unternehmen selbst bei eindeutiger Branchenzugehörigkeit in bestimmten jahresabschlussrelevanten Eigenschaften erheblich voneinander unterscheiden können.<sup>6</sup> So könnten die einer bestimmten Branche angehörigen Unternehmen ihren Jahresabschluss auf der Grundlage unterschiedlicher Rechnungslegungsnormen erstellen. Aber auch bei der Verwendung gleicher Normen könnten sich eingeräumte Bilanzierungswahlrechte und Ermessensspielräume erheblich auf die Vergleichbarkeit von Unternehmen auswirken.<sup>7</sup> Geographische Unterschiede könnten sich nachhaltig auf die Kostenstrukturen eines Unternehmens in Bezug auf Transport und Produktion auswirken.<sup>8</sup> Weitere Differenzierungsmerkmale könnten in den Kapitalverhältnissen, unternehmensexternen Einflussfaktoren auf Bilanz und Gewinn- und Verlustrechnung (wie etwa staatlichen Regulierungen), Kapazitätsauslastungsgraden, in der Produktionstechnologie, dem Alter und der Struktur des Anlagevermögens und der Kundenstrukturen begründet sein.<sup>9</sup> Nicht zu vernachlässigen ist in diesem Zusammenhang auch die Konzernzugehörigkeit eines Unternehmens. Ein ver-

---

<sup>1</sup> Vgl. MCKEE (1989), S. 109.

<sup>2</sup> Vgl. GAUNTT/GLEZEN (1997), S. 57.

<sup>3</sup> Vgl. LEE/COLBERT (1997), S. 395.

<sup>4</sup> Vgl. MÜLLER (1996), S. 53.

<sup>5</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 210; BLOCHER/WILLINGHAM (1985), S. 88.

<sup>6</sup> Vgl. PEEMÖLLER (1992), S. Sp. 347-348; SPERL (1978), S. 278.

<sup>7</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 10.

<sup>8</sup> Vgl. MCKEE (1989), S. 111.

<sup>9</sup> Vgl. GÄRTNER (1994), S. 53; LEE/COLBERT (1997), S. 395.

bundenes Unternehmen könnte aufgrund der Leistungsverflechtung innerhalb des Konzerns erhebliche Unterschiede im Bereich der Forderungen aufweisen oder wesentlich günstigere Einkaufspreise erhalten als ein nicht zu einem Konzern angehörendes Unternehmen.<sup>1</sup>

Eine Möglichkeit, derartigen Problemen teilweise entgegenzuwirken, liegt darin, dass der Prüfer einen eigenen Durchschnittswert ermittelt und nicht auf einen bereits aufbereiteten Durchschnittswert zurückgreift, der zumeist auch eine Vielzahl von Unternehmen umfasst, die aus o.a. Gründen weniger mit dem geprüften Unternehmen verglichen werden können.<sup>2</sup> Die hierzu notwendigen Daten könnte der Prüfer aus bereits von ihm oder seiner Prüfungsgesellschaft geprüften Mandanten erhalten oder aus Datenbanken (etwa HOPPENSTEDT<sup>3</sup>), die eine Vielzahl von geprüften Jahresabschlüssen umfassen. Zwar würde die Ermittlung eines eigenen Branchendurchschnittswerts aufgrund von Schwierigkeiten bei der Identifizierung von geeigneten Unternehmen oder der Beschaffung der Jahresabschlüsse durchaus viel Zeit und Kosten verursachen.<sup>4</sup> Allerdings bietet diese Vorgehensweise auch wesentliche Vorteile gegenüber der Verwendung von herkömmlichen Branchendurchschnittswerten, da der Prüfer eine eigene Auswahl trifft, in der er ausschließlich solche Unternehmen berücksichtigen kann, welche die wesentlichen Eigenschaften des Mandanten auch wirklich teilen. Darüber hinaus könnte er den einbezogenen Unternehmen sogar unterschiedliche Gewichtungen zuordnen, um auf diesem Wege eine noch höhere Vergleichbarkeit zu erreichen.<sup>5</sup>

#### **2.2.3.1.3.2.2 Unternehmens-Benchmarking**

Bei der zweiten Art des Branchen-Benchmarkings, dem sog. Unternehmens-Benchmarking, greift der Prüfer nicht auf Durchschnittswerte zurück, die aus den Daten mehrerer Unternehmen ermittelt wurden, sondern auf Daten eines einzelnen, dem Mandanten möglichst vergleichbaren Unternehmens.<sup>6</sup> Anders als bei der zuvor dargestellten Ermittlung eines eigenen Branchendurchschnitts fließen die Daten des Unternehmens

---

<sup>1</sup> Vgl. MCKEE (1989), S.111.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 89; MCKEE (1989), S. 111.

<sup>3</sup> Weitere Informationen zur Jahresabschlusssdatenbank finden sich unter <http://www.hoppenstedt.de> (Stand: 28.11.2006).

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 89.

<sup>5</sup> Vgl. MCKEE (1989), S. 109.

<sup>6</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 308.

hier folglich ungewichtet in die Vergleichsgröße ein. Daher ist der Identifikation eines Benchmark-Unternehmens hierbei auch eine wesentlich höhere Bedeutung zuzuordnen.

Gerade hierin liegt jedoch auch die größte Schwierigkeit des Unternehmens-Benchmarkings, denn der Prüfer ist bei der Suche auf ein angemessenes Maß an Kenntnissen über das einzelne Unternehmen angewiesen.<sup>1</sup> Unternehmen sind allerdings häufig bemüht, ihre internen Daten möglichst geheim zu halten, da sie den Verlust ihrer Wettbewerbsvorteile befürchten.<sup>2</sup> Daher scheint die Identifikation eines Benchmark-Unternehmens wenn überhaupt nur unter den eigenen Mandanten möglich.<sup>3</sup> Bei nicht selbst geprüften Unternehmen dürfte sich diese Selektion als sehr schwierig oder sogar kaum durchführbar erweisen.

#### **2.2.3.1.3.3 Vorteile und Nachteile des Branchen-Benchmarkings**

Die Vorteile des Branchen-Benchmarkings liegen darin, dass die herangezogenen Vergleichsinformationen aufgrund der fehlenden Beeinflussbarkeit durch das geprüfte Unternehmen generell durch eine höhere Zuverlässigkeit gekennzeichnet sind als interne Informationen.<sup>4</sup> Allerdings sollte der Prüfer den Grad der Zuverlässigkeit auch davon abhängig erachten, ob die verwendeten Branchenwerte auf Daten zurückzuführen sind, die zuvor einer Jahresabschlussprüfung unterzogen wurden oder nicht.<sup>5</sup> Darüber hinaus ist beim Branchen-Benchmarking ebenfalls positiv zu berücksichtigen, dass hierbei Werte der gleichen Periode gegenübergestellt werden, denen gleiche gesamtwirtschaftliche und branchenspezifische Bedingungen zugrunde liegen.<sup>6</sup>

Als ein grundsätzlicher Nachteil des Branchen-Benchmarkings ist aufzuführen, dass auf Branchenwerte für das zu prüfende Geschäftsjahr nur dann zurückgegriffen werden kann, wenn die betrachtete Jahresabschlussprüfung erst relativ spät in der Prüfungssaison erfolgt, da vorher noch nicht damit zu rechnen ist, dass aktuelle und geprüfte Branchenwerte vorliegen.<sup>1</sup> Da jedoch auch Unternehmen existieren, bei denen Rumpfgeschäftsjahre vorliegen, können Prüfer ersatzweise auf Branchenwerte zurückgreifen, die im günstigsten Fall nur wenige Monate vor dem Bilanzstichtag des zu prüfenden Unter-

---

<sup>1</sup> Vgl. KNECHEL (2001), S. 282.

<sup>2</sup> Vgl. KNECHEL (2001), S. 282.

<sup>3</sup> Vgl. MÜLLER (1996), S. 53.

<sup>4</sup> Vgl. MÜLLER (1996), S. 53; MCKEE (1989), S. 103.

<sup>5</sup> Vgl. MCKEE (1989), S. 109.

<sup>6</sup> Vgl. CASHIN/NEUWIRTH/LEVY (1986), S. 41-8.

nehmens datiert sind. Allerdings sollte der Prüfer in Erfahrung bringen, ob nicht kürzlich eingetretene und möglicherweise unberücksichtigte Ereignisse und Änderungen in der Branche oder Gesamtwirtschaft die Vergleichbarkeit der Perioden beeinträchtigen.<sup>2</sup>

#### **2.2.3.1.4 Betriebs-Benchmarking**

Eine Sonderform des Benchmarkings ist das Betriebs-Benchmarking. Hierbei wird nicht auf Daten eines externen Unternehmens zurückgegriffen, sondern auf Informationen eines zum betrachteten Unternehmen gehörenden Betriebs oder Unternehmensbereichs.<sup>3</sup> Daher spricht man in diesem Fall auch häufig vom *internen Benchmarking*. Der wesentliche Vorteil gegenüber den zuvor dargestellten externen Vergleichen liegt in der besseren Verfügbarkeit und Zugänglichkeit der relevanten Daten.

#### **2.2.3.2 Einfache Rechenmodelle**

##### **2.2.3.2.1 Wesen einfacher Rechenmodelle**

Einfache Rechenmodelle sind „nicht-statistische“ modellbasierte Verfahren, die mit einfachen mathematischen Operationen und wenigen Daten auskommen.<sup>4</sup> Einfache Rechenmodelle stellen überschlägige Berechnungen dar, bei denen zur Ermittlung eines Erwartungswerts sachlogische Verknüpfungen zwischen der zu prüfenden Jahresabschlussgröße und sonstigen Daten genutzt werden.<sup>5</sup> Hierbei ist grundsätzlich zwischen Modellen, die nur eine unabhängige Variable berücksichtigen, und Modellen, die auf mehrere unabhängige Variablen zurückgreifen, zu unterscheiden. Es wird angenommen, dass die Aussagekraft des Modells und somit die Verlässlichkeit des gebildeten Erwartungswerts mit der Anzahl der berücksichtigten unabhängigen Variablen zunimmt.<sup>1</sup> Allerdings steigt mit zunehmender Zahl der berücksichtigten Einflussfaktoren auch die Komplexität des Modells und damit auch der benötigte Zeitaufwand, was ihrer Effizienz

---

<sup>1</sup> Ähnlich argumentieren GAUNTT/GLEZEN (1997), S. 57.

<sup>2</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 6-7.

<sup>3</sup> Handelt es sich bei dem betrachteten Unternehmen um ein Mutter- oder Konzernunternehmen, kommt als Vergleichsobjekt auch ein eigenständiges Tochterunternehmen in Betracht.

<sup>4</sup> Ein weiteres modellbasiertes Verfahren stellt die Regressionsanalyse dar. Im Gegensatz zu einfachen Rechenmodellen kommen Regressionsanalysen i.d.R. nicht ohne IT-gestützte Prüfungstechniken aus. Da es sich bei Regressionsanalysen per definitionem jedoch um Verfahren handelt, die neben Daten der zu prüfenden Periode auch Daten aus der Vergangenheit benötigen, bilden diese Verfahren den Gegenstand des Kap. 2.2.4.

<sup>5</sup> Vgl. ALDERMAN (1987), S. 92.

wiederum entgegenwirkt.<sup>2</sup> Darüber hinaus argumentierten BLOCHER/WILLINGHAM, dass die Verwendung lediglich einer unabhängigen Variablen auch deshalb vorzuziehen sei, weil sich die Wirkungen der anderen nicht berücksichtigten Einflussfaktoren häufig gegenseitig aufheben würden, so dass bei Modellen mit nur einer unabhängigen Variablen nicht mit erheblichen Qualitätsverlusten zu rechnen sei.<sup>3</sup> Dennoch sollte die berücksichtigte Anzahl von beeinflussenden Variablen vom Einzelfall abhängig gemacht werden. Sicherlich wäre dabei auch erneut die Verfügbarkeit und Verlässlichkeit der Daten zu berücksichtigen. Letztendlich kommt es allerdings entscheidend auf die Fähigkeiten des Prüfers an, die wesentlichen Einflussfaktoren zu erkennen und mögliche überflüssige, sich gegenseitig kompensierende Faktoren unberücksichtigt zu lassen.

#### **2.2.3.2.2 Anwendungsbereiche von einfachen Rechenmodellen**

Der Einsatz einfacher Rechenmodelle kommt insbesondere zur Ermittlung von Erwartungswerten im Bereich der Gewinn- und Verlustrechnung in Betracht. Da die Bestandskonten der Bilanz häufig die Ereignisse mehrerer Perioden umfassen und ihre Stichtagsbetrachtung häufig zufällige Schwankungen zur Folge hat, eignen sie sich weniger für die Anwendung von einfachen Rechenmodellen.<sup>4</sup> Obwohl dem Prüfer zur Ermittlung eines Erwartungswerts eine Vielzahl von möglichen sonstigen Daten, wie etwa andere Jahresabschlussgrößen<sup>5</sup> oder unternehmensexterne Daten<sup>6</sup> zur Verfügung stehen, greifen einfache Rechenmodelle in der Regel auf Betriebsdaten, also von der Finanzbuchführung unabhängige Größen zurück.<sup>7</sup> Dieses ist darauf zurückzuführen, dass Betriebsdaten in der Regel als Mittel zur internen Steuerung und Kontrolle des operativen Geschäfts herangezogen werden und daher einen direkten Bezug zu den in der Gewinn- und Verlustrechnung ausgewiesenen Ergebnissen der operativen Tätigkeit des Unternehmens aufweisen.<sup>8</sup>

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 108.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 102.

<sup>3</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 103.

<sup>4</sup> Vgl. WILSON/COLBERT (1991), S. 40. Einige wenige Beispiele für den Einsatz von einfachen Rechenmodellen in der Bilanz gibt McKEE (1989), S. 120.

<sup>5</sup> Vgl. GAUNTT/GLEZEN (1997), S. 57; FITZSIMONS/LEVINE (1989), S. 8; McKEE (1989), S. 114; BLOCHER/WILLINGHAM (1985), S. 100.

<sup>6</sup> Vgl. SCOTT/WALLACE (1994), S. 65; WALLACE (1983a), S. 32; BLOCHER/PATTERSON (1996), S. 54; KINNEY/FELIX (1980), S. 101.

<sup>7</sup> Vgl. BLOCHER (1983), S. 28; WILSON/COLBERT (1991), S. 40; MÜLLER (1996), S. 36-37.

<sup>8</sup> Vgl. McKEE (1989), S. 117.



Beispiele für einfache Rechenmodelle zur Beurteilung von Konten der Gewinn- und Verlustrechnung existieren in der Literatur in unüberschaubarer Fülle und können für die unterschiedlichsten Branchen wiedergegeben werden.<sup>1</sup> Die Möglichkeiten zur Modellierung von sachlogischen Zusammenhängen zwischen einzelnen Erfolgskonten und den relevanten Betriebsdaten sind dabei vielfältig und letztendlich von der Kreativität und den Kenntnissen des Prüfers über den Mandanten und der Branche abhängig.<sup>2</sup> So können beispielsweise die Umsatzerlöse anhand der verkauften Mengen und des durchschnittlichen Verkaufspreises je Produktgruppe<sup>3</sup> oder die Umsatzkosten anhand der produzierten Mengen und den durchschnittlichen Herstellungskosten<sup>4</sup> plausibilisiert werden. Der Zinsaufwand kann anhand der durchschnittlichen Darlehenssumme und dem durchschnittlichen Marktzins ebenso geeignet überschlägig berechnet werden wie der Personalaufwand unter Zuhilfenahme der Mitarbeiterzahlen und -löhne oder Reparaturaufwendungen anhand der Maschinenstundenzahl und dem Alter der Anlagen.<sup>5</sup> Bei Taxi- oder Speditionsbetrieben lassen sich die Kraftstoffkosten mittels der Fahrleistung der Periode und dem durchschnittlichen Verbrauch je 100 km errechnen. Ebenso können die Umsatzerlöse von Hotelunternehmen anhand der Anzahl der Zimmer multipliziert mit der Belegungsrate und den Übernachtungspreisen ermittelt werden.<sup>6</sup>

#### **2.2.3.2.3 Vorteile und Nachteile von Rechenmodellen**

Ein wesentlicher Vorteil von Rechenmodellen liegt darin, dass der Prüfer zur Bildung eines Erwartungswerts *explizieren* muss, auf welche unabhängige Variable(n) er zurückgreift und wie sich diese auf die zu prüfende Variable auswirken.<sup>7</sup> Bei den bisher vorgestellten Verfahren analytischer Prüfungshandlungen war die Erwartungsbildung demgegenüber ausschließlich *implizit*,<sup>8</sup> da sie davon ausgehen, dass die Normenkonformität der betrachteten Jahresabschlussgröße oder Kennzahl durch einen bloßen Vergleich mit Werten aus der Vergangenheit bzw. mit Werten von anderen Unternehmen oder mit den Planzahlen des zu prüfenden Unternehmens abgeleitet werden kann. Unerwartete oder unvorhersehbare Geschehnisse im Unternehmen oder dessen Umfeld

---

<sup>1</sup> Eine branchenabhängige Übersicht gibt etwa SMITH (1989), S. 65 und LEE/COLBERT (1997), S. 396.

<sup>2</sup> Vgl. MÜLLER (1996), S. 38; GÄRTNER (1994), S. 107.

<sup>3</sup> Vgl. IDW (2006), R 385.

<sup>4</sup> Vgl. FITZSIMONS/LEVINE (1989), S. 8-9.

<sup>5</sup> Vgl. MCKEE (1989), S. 117.

<sup>6</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 309.

<sup>7</sup> Vgl. SCOTT/WALLACE (1994), S. 65.

<sup>8</sup> Vgl. SCOTT/WALLACE (1994), S. 65.

sind in diesen Verfahren in der Regel jedoch nicht berücksichtigt, was ihrer Verlässlichkeit in einer dynamischen Umwelt z.T. entgegenwirkt. Modellbasierte Verfahren hingegen können derartige Ereignisse explizit berücksichtigen und können daher zu stärkeren Prüfungsnachweisen beitragen.<sup>1</sup>

Des Weiteren bringt die Anwendung modellbasierter Verfahren mit sich, dass der Prüfer seine Vorgehensweise bei der Ermittlung des Erwartungswerts strukturiert und somit die Wahrscheinlichkeit verringert wird, dass der Prüfer wesentliche Einflussgrößen oder deren Veränderung im zu prüfenden Geschäftsjahr übersieht und folglich falsche Schlussfolgerungen aus festgestellten Abweichungen bzw. Übereinstimmungen ableitet. So könnte der Prüfer beispielsweise bei der Untersuchung der Kraftstoffaufwendungen eines Speditionsunternehmens davon ausgehen, dass eine Steigerung zum Vorjahr um etwa 6% im Einklang mit der Erhöhung der Umsatzerlöse steht. Bei einer entsprechenden Durchführung eines Rechenmodells würde der Prüfer hingegen relevante Faktoren, wie etwa die Anzahl der LKW, die gefahrenen Kilometer oder die Treibstoffpreise hinzuziehen. Hierbei könnte dem Prüfer auffallen, dass der Mandant zur Senkung seiner Kosten den Bestand der LKW im abgelaufenen Geschäftsjahr um 25% reduziert hat. Unter diesen Umständen würde er die Erhöhung der ausgewiesenen Kraftstoffaufwendungen sicherlich in einem anderen Lichte betrachten.<sup>2</sup>

Ähnlich könnte es sich bei der Beurteilung der Fertigungslöhne eines Industrieunternehmens darstellen.<sup>3</sup> Sollte der Prüfer hierbei etwa eine Erhöhung um 9-10% feststellen, könnte er davon ausgehen, dass dies auf die durchschnittliche Erhöhung des Gehaltneiveaus von 8-12% zurückzuführen ist. Zur überschlägigen Berechnung mittels eines einfachen Rechenmodells würde der Prüfer allerdings auch die Mitarbeiterzahlen und die gesamten Arbeitsstunden im abgelaufenen Geschäftsjahr in Betracht ziehen. Dabei könnte ihm auffallen, dass die Arbeitsstunden aufgrund eines technisch bedingten Produktionsausfalls in einem bestimmten Zeitraum erheblich geringer waren als üblich. In einer derartigen Situation würde der Prüfer die Steigerung der Fertigungslöhne sicherlich anders beurteilen als bei der bloßen Vornahme eines Vorjahresvergleichs oder einer Trendanalyse.

Die Durchführung einfacher Rechenmodelle und das damit verbundene Berücksichtigen von relevanten Einflussfaktoren kann also dazu führen, dass der Prüfer auf Sachverhalte

---

<sup>1</sup> Vgl. WALLACE (1983a), S. 32.

<sup>2</sup> Vgl. MCKEE (1989), S. 116.

<sup>3</sup> Vgl. im Folgenden BLOCHER/WILLINGHAM (1985), S. 101.

aufmerksam wird, die von wesentlicher Bedeutung für die Beurteilung der Angemessenheit der zu prüfenden Jahresabschlussgröße sind. Dieses ermöglicht dem Prüfer, wesentlich effektiver vorzugehen als bei der Anwendung von analytischen Prüfungshandlungen, die lediglich eine implizite Erwartungsbildung umfassen.<sup>1</sup>

Neben einer expliziten Erwartungsbildung weisen einfache modellbasierte Verfahren weiterhin den Vorteil auf, dass sie die zu prüfenden Jahresabschlussgrößen vor dem Hintergrund der relevanten Betriebsdaten betrachten.<sup>2</sup> Diese Vorgehensweise eignet sich häufig besser als die bloße Zuhilfenahme von Jahresabschlussdaten.<sup>3</sup> Mithin liegt der Grund für Änderungen von Jahresabschlussgrößen häufig in betrieblichen Ereignissen, so dass Rechenmodelle die Möglichkeit bieten, zentrale Einflussfaktoren mit einzubeziehen.<sup>4</sup>

Die Verwendung von Betriebsdaten ist auch deshalb positiv zu bewerten, weil ihre Herkunft häufig unabhängig von der für das externe Rechnungswesen verantwortlichen Abteilung ist (etwa aus der Marketing- oder Produktionsabteilung<sup>5</sup>), und sie somit verlässlicher sind.<sup>6</sup> Zudem sind betriebliche Daten aufgrund ihrer Zugehörigkeit zur gleichen Periode wie die zu prüfenden Daten häufig einfacher zu beschaffen als Daten, die der Vorperiode oder einem davor liegenden Zeitraum zuzuordnen sind.<sup>7</sup>

Ein wesentliches Problem von einfachen modellbasierten Verfahren hängt mit ihrer Verwendung von betrieblichen Daten zusammen. Die Verfügbarkeit von jeglichen Betriebsdaten kann nicht als selbstverständlich vorausgesetzt werden und ist z.T. von der Bereitschaft des Mandanten abhängig, diese dem Prüfer zu überlassen. Zwar stellt die Kostenrechnung und damit eine Vielzahl von betrieblichen Daten einen wesentlichen Bestandteil der Jahresabschlusserstellung und -prüfung dar,<sup>8</sup> allerdings besteht seitens der Unternehmen auch das Interesse, ihre betrieblichen Zahlen geheim zu halten, da sie möglicherweise sensible Informationen (etwa Kosten- und Wettbewerbsvorteile) beinhalten. Der Prüfer könnte hier versuchen, die relevanten Daten selbst aufzubereiten. Da allerdings davon auszugehen ist, dass die nachträgliche Rekonstruktion von Betriebsdaten vergleichsweise mehr Zeit in Anspruch nehmen und höhere Kosten verursachen

---

<sup>1</sup> Vgl. MCKEE (1989), S. 115.

<sup>2</sup> Vgl. BLOCHER/PATTERSON (1996), S. 54.

<sup>3</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 309; SCOTT/WALLACE (1994), S. 65.

<sup>4</sup> Vgl. BLOCHER/PATTERSON (1996), S. 54.

<sup>5</sup> Vgl. WALLACE (1983a), S. 32.

<sup>6</sup> Vgl. GAUNTT/GLEZEN (1997), S. 57.

<sup>7</sup> Vgl. ALDERMAN (1987), S. 93.

<sup>8</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 488-491.

würde, sollte von diesem Verfahren abgesehen und stattdessen ein anderes kostengünstigeres Verfahren eingesetzt werden.<sup>1</sup>

Neben der Verfügbarkeit von Betriebsdaten ist auch ihre Verlässlichkeit genau zu hinterfragen.<sup>2</sup> Betriebsdaten stellen in der Regel ungeprüfte Daten dar.<sup>3</sup> Darüber hinaus unterliegen sie häufig weniger strengen internen Kontrollen des Unternehmens als die für den Jahresabschluss relevanten Finanzdaten.<sup>4</sup> Der Prüfer sollte also genau zurück verfolgen, woher die bereitgestellten Daten kommen und welche Unsicherheiten damit verbunden sind.<sup>5</sup> Nur bei Systemen, die ausreichenden und regelmäßigen Qualitätskontrollen unterzogen werden, kann sich der Prüfer darauf verlassen, dass ihm vertrauenswürdige Daten zur Verfügung stehen, die zu verlässlichen Prüfungsnachweisen führen.<sup>6</sup>

Ein weiterer wesentlicher Nachteil von einfachen modellbasierten Verfahren ist in ihrem stark subjektiven Charakter zu sehen.<sup>7</sup> Trotz eines gewissen Formalisierungsgrads ist die Zuverlässigkeit von Rechenmodellen nicht unwesentlich von den Fähigkeiten des jeweiligen Prüfers abhängig. Da es entscheidend darauf ankommt, dass der Prüfer die wesentlichen Einflussfaktoren der zu prüfenden Jahresabschlussgröße identifiziert, wird von ihm ein umfangreiches Wissen über die Geschäftsprozesse des Mandanten und den Industriezweig verlangt.<sup>8</sup> Darüber hinaus obliegt es dem Prüfer, die identifizierten Variablen in der richtigen Form miteinander zu verknüpfen.<sup>9</sup> Diese Fähigkeiten erlangt der Prüfer allerdings erst nach einer bestimmten Zeit, so dass die Anwendbarkeit und Zuverlässigkeit von Rechenmodellen wesentlich von der Erfahrung des Prüfers abhängt und ihre Ergebnisse unterschiedlich interpretiert werden können.<sup>10</sup>

---

<sup>1</sup> Vgl. MCKEE (1989), S. 122.

<sup>2</sup> Vgl. ARENS/ELDER/BEASLEY (2005), S. 215.

<sup>3</sup> Vgl. WALLACE (1983a), S. 32.

<sup>4</sup> Vgl. MCKEE (1989), S. 122.

<sup>5</sup> Vgl. IDW (2006), R 388.

<sup>6</sup> Vgl. LEE/COLBERT (1997), S. 397.

<sup>7</sup> Vgl. GÄRTNER (1994), S. 109.

<sup>8</sup> Vgl. IDW (2006), R 386.

<sup>9</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 109.

<sup>10</sup> Vgl. IDW (2006), R 386; BLOCHER/WILLINGHAM (1985), S. 109.

## **2.2.4 Informationen aus der Vergangenheit und aus dem aktuellen Geschäftsjahr**

### **2.2.4.1 Einordnung mathematisch-statistischer Verfahren**

Eine Sonderstellung bei den Verfahren analytischer Prüfungshandlungen nehmen die mathematisch-statistischen Verfahren ein, da sie zur Entwicklung eines Erwartungswerts sowohl auf Daten vergangener Geschäftsjahre als auch auf Daten aus der zu prüfenden Periode zurückgreifen. Als mathematisch-statistische Verfahren bei der Durchführung von analytischen Prüfungshandlungen kommen im Wesentlichen Regressionsanalysen und der Box-Jenkins-Ansatz in Betracht.<sup>1</sup> Darüber hinaus werden komplexere Verfahren, wie Martingale und X-11, diskutiert.<sup>2</sup> Da es im Rahmen dieser Untersuchung nicht darum geht, eine ausführliche Darstellung der unterschiedlichen mathematisch-statistischen Verfahren vorzunehmen, erfolgt im Folgenden eine exemplarische Darstellung anhand der Regressionsanalyse.

### **2.2.4.2 Die Regressionsanalyse als beispielhafte Darstellung eines mathematisch-statistischen Verfahrens**

#### **2.2.4.2.1 Begriffsbestimmung und Wesen von Regressionsanalysen**

Unter dem Begriff der Regression ist die stochastische Abhängigkeit zwischen einer abhängigen Variable und einer oder mehreren unabhängigen Variablen zu verstehen.<sup>3</sup> Stochastische Abhängigkeiten stellen dabei Zusammenhänge zwischen Variablen (etwa Jahresabschlussdaten) dar, die nicht streng funktional sind und daher nicht eindeutig bestimmt werden können.<sup>4</sup> Dieses äußert sich in der Regel darin, dass eine zufallsbedingte Streuung der Werte einer abhängigen Variable vorliegt und ein bestimmter Wert nicht mit Sicherheit erwartet werden kann.<sup>5</sup> Gründe für lediglich stochastische Abhängigkeiten sind beispielsweise darin zusehen, dass die abhängige Variable von weiteren Fakto-

---

<sup>1</sup> Vgl. BOX/JENKINS (1976). Eine ausführliche Darstellung der Einsetzbarkeit im Rahmen von analytischen Prüfungshandlungen findet sich bei GÄRTNER (1994), S. 155-161.

<sup>2</sup> Vgl. CHEN/LEITCH (1999), S. 35-69. Eine kurze Beschreibung beider Verfahren findet sich bei LEE/COLBERT (1997), S. 396-397.

<sup>3</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 46.

<sup>4</sup> Vgl. SPERL (1978), S. 220.

<sup>5</sup> Vgl. RÖNZ/FÖRSTER (1992), S. 5.

ren beeinflusst wird, die jedoch nicht berücksichtigt werden, sowie in der ungenauen Messung von Beobachtungen oder in rein zufälligen Einflüssen.<sup>1</sup>

Ein Gebiet, bei dem zu vermuten ist, dass zwar möglicherweise funktionale Zusammenhänge zwischen den einzelnen Jahresabschlussdaten vorliegen, aufgrund der in der Regel vorherrschenden Datenkomplexität allerdings eindeutige und feste Beziehungen nur sehr schwierig auszumachen sind, stellen Jahresabschlüsse dar.<sup>2</sup> Um dennoch einen zuverlässigen Erwartungswert zur Durchführung von analytischen Prüfungshandlungen generieren zu können, kann der Prüfer entweder anhand seiner bisherigen Erfahrung bestimmte funktionale Zusammenhänge schätzen. Dieses in Kap. 2.2.3.2 als „nicht-statistisches“ modelbasiertes Verfahren vorgestellte Vorgehen ist allerdings stark subjektiv geprägt, da es von den individuellen Erfahrungen und Fähigkeiten des Prüfers abhängt.<sup>3</sup> Eine andere und weitaus objektivere Möglichkeit besteht darin, die stochastischen Abhängigkeiten zwischen der abhängigen und einer oder mehreren unabhängigen Variablen durch die Anwendung einer Regressionsanalyse zu approximieren.<sup>4</sup> Die sich hieraus ergebende Funktion wird im Gegensatz zu einer streng mathematischen Funktion als Regressionsfunktion bezeichnet.<sup>5</sup> Diese ist gewissermaßen als eine mathematische Formel zu verstehen, auf deren Grundlage der Erwartungswert für die zu prüfende Jahresabschlussgröße unter Rückgriff auf die Beobachtungswerte der Vergangenheit prognostiziert wird.<sup>6</sup> Dieses hat gegenüber den „nicht-statistischen“ modelbasierten Verfahren den Vorteil, dass die Erwartungsbildung nicht vom intersubjektiv differierenden Wissen und Können abhängig gemacht wird, sondern „the statistical techniques [...] let the data ,speak for itself“<sup>7</sup>.

#### **2.2.4.2.2 Formen von Regressionsanalysen**

Generell können Regressionsanalysen danach unterschieden werden, ob zur Bildung eines Erwartungswerts auf die stochastischen Zusammenhänge zwischen einer unabhängigen Variablen oder mehreren unabhängigen Variablen und der abhängigen Größe zu-

---

<sup>1</sup> Vgl. GÄRTNER (1994), S. 138.

<sup>2</sup> Vgl. KASK (1979), S. 35.

<sup>3</sup> Für die Bedeutung von Erfahrung siehe etwa QUICK (1999), S. 214.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 305-306.

<sup>5</sup> Vgl. RÖNZ/FÖRSTER (1992), S. 5.

<sup>6</sup> Vgl. STEWARD (1978), S. 70.

<sup>7</sup> Vgl. MCKEE (1989), S. 124.

rückgegriffen wird.<sup>1</sup> Erstgenannte werden generell als Einfach-Regressionsanalysen bezeichnet und Letztgenannte hingegen als Mehrfach- oder multiple Regressionsanalysen.<sup>2, 3</sup>

Als Beispiel für die Durchführung von analytischen Prüfungshandlungen anhand einer Einfach-Regressionsanalyse könnte die Beurteilung der im Jahresabschluss ausgewiesenen Forderungen aus Lieferungen und Leistungen (im Folgenden als Forderungen aLuL bezeichnet) herangezogen werden, wobei die Umsatzerlöse der entsprechenden Periode als die den Forderungsbestand beeinflussende Variable zu betrachten ist. Zur Ermittlung der Regressionsfunktion unterstellt der Prüfer häufig einen zumindest linearen Zusammenhang,<sup>4</sup> so dass es sich in diesem Fall um eine *lineare Einfachregression* handelt.<sup>5</sup> Zur graphischen Darstellung dieser Vorgehensweise sind zunächst die Beobachtungswerte aus der Vergangenheit in ein sog. Streuungsdiagramm zu übertragen. In die sich hierbei ergebende Punktwolke ist anschließend eine Gerade einzuzeichnen, die sich den Datenpunkten möglichst gut anpasst (Abbildung 2).<sup>6</sup>

---

<sup>1</sup> Vgl. LEE/COLBERT (1997), S. 396.

<sup>2</sup> Vgl. MCKEE (1982), S. 40-41.

<sup>3</sup> Einen Spezialfall der Einfach-Regression stellt die Zeitreihenanalyse dar, die die Abhängigkeit einer Variablen von der Zeit untersucht. In den Bereich der Zeitreihenanalysen fallen insbesondere die Trendanalysen (siehe hierzu auch Kap. 2.2.2.2). Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 49.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 305-306 sowie GÄRTNER (1994), S. 139. Für diese vereinfachende Annahme können eine Vielzahl von Argumenten hervorgebracht werden. So können etwa viele Ursache-Wirkungs-Zusammenhänge zumindest annähernd in eine lineare Beziehung zueinander gesetzt werden. Darüber hinaus lassen sich auch nicht-lineare Beziehungen häufig ohne großen Aufwand in eine lineare Beziehung transformieren. Zudem stellen die den linearen Beziehungen zugrunde liegenden Geradengleichungen formal betrachtet die einfachste Alternative dar, einen Ursache-Wirkungszusammenhang zweier Variablen darzustellen. Und letztendlich sind Geradengleichungen einfach zu interpretieren. Siehe hierzu SCHLITTGEN (2003), S. 414.

<sup>5</sup> Vgl. KLITZSCH/HELLMUND/SCHUMANN (1992), S. 188-189.

<sup>6</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 56.

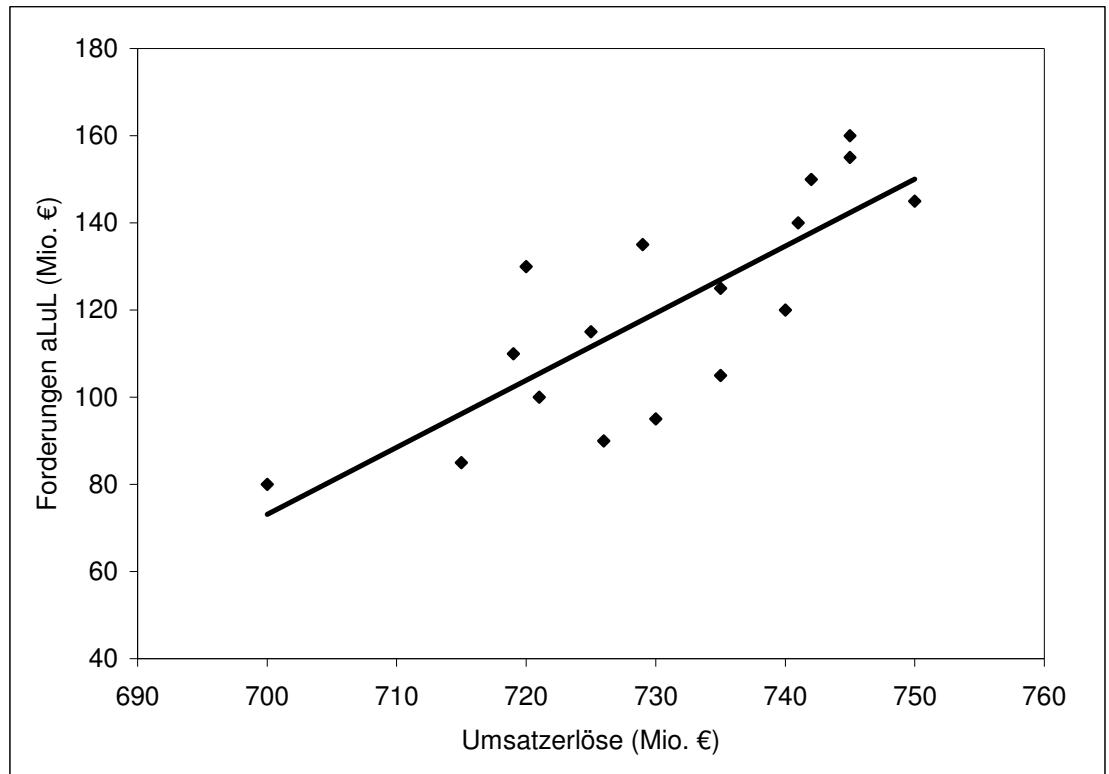


Abbildung 2: Regressionsgerade zur Beurteilung der ausgewiesenen Forderungen aLuL anhand der Umsatzerlöse

Da theoretisch eine Vielzahl von Geraden durch die Punktwolke gelegt werden könnten, besteht das Ziel der linearen Einfachregression darin, die Lage derjenigen Geraden zu bestimmen, die den Zusammenhang zwischen der unabhängigen und der abhängigen Größe best möglichst approximiert.<sup>1</sup> Die Lage dieser sog. Regressionsgeraden wird im Wesentlichen durch zwei Parameter bestimmt. Dabei handelt es sich einerseits um das konstante Glied  $a$ , welches den Schnittpunkt der Geraden mit der Ordinate angibt, und andererseits um den Regressionskoeffizienten  $b$ , welcher die Steigung der Geraden darstellt. Für die Funktion der Regressionsgeraden ergibt sich entsprechend folgende Notation:<sup>2</sup>

$$Y = a + bX$$

<sup>1</sup> Vgl. SCHLITTGEN (2003), S. 416.

<sup>2</sup> Vgl. SCHLITTGEN (2003), S. 416.



mit

$Y$  = abhängige Variable (hier der Erwartungswert für die zum Geschäftsjahresende ausgewiesenen Forderungen aLuL)

$X$  = erklärende Variable (hier die Umsatzerlöse des betrachteten Geschäftsjahrs)

$a$  = absolutes Glied

$b$  = Regressionskoeffizient

Wie Abbildung 2 zu erkennen gibt, ist es unmöglich, eine Regressionsgerade zu bestimmen, die durch alle Beobachtungspunkte verläuft. In der Regel liegen die Beobachtungswerte über- oder unterhalb der Regressionsgeraden. Dieses ist generell darauf zurückzuführen, dass es sich bei der Regressionsanalyse um die formale Darstellung von stochastischen und nicht deterministischen Zusammenhängen handelt.<sup>1</sup> Die den stochastischen Abhängigkeiten immanente zufallsbedingte Streuung könnte formal dadurch berücksichtigt werden, dass die o.a. Regressionsfunktion durch eine sog. Störvariable  $e$  erweitert wird, die wie folgt berechnet werden kann:<sup>2</sup>

$$e_i = y_i - y_i^l$$

mit

$y_i$  = Beobachtungswert der abhängigen Variablen für  $x_i$

$y_i^l$  = Schätzwert der abhängigen Variable für  $x_i$  (durch Regression ermittelt)

$e_i$  = nicht erklärbare Abweichung zwischen Beobachtungs- und Schätzwert

Nimmt man diese auch als *Residuum* bezeichnete Variable  $e$  zusätzlich in die bisherige Formel der Regressionsgeraden auf, ergibt sich entsprechend folgende Regressionsgleichung:<sup>3</sup>

$$Y = a + bX + e$$

Die Gleichung umfasst somit eine systematische Komponente ( $a + bX$ ), die sich in einem stabilen Verhältnis zur erklärenden Variable ändert, und eine nicht erklärbare sto-

---

<sup>1</sup> So auch GÄRTNER (1994), S. 137-138.

<sup>2</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 57.

<sup>3</sup> Vgl. SCHLITTGEN (2003), S. 423-424. GÄRTNER (1994) ist hingegen der Auffassung, dass der Störvariable bei der Anwendung im Rahmen der Jahresabschlussprüfung keine große Bedeutung beizumessen sei, so dass ihre Nicht-Berücksichtigung die Eignung der Regressionsanalyse zum Zwecke analytischer Prüfungshandlungen nicht erheblich beeinträchtigen würde. Siehe GÄRTNER (1994), S. 140.

chastische Zufallskomponente  $e$ . Zum Zwecke der Ermittlung eines zuverlässigen Erwartungswerts wäre es von besonderem Nutzen, wenn die Zufallskomponente  $e$  gleich null oder zumindest möglichst gering wäre. Da dieses allerdings für den Bereich des Jahresabschlusses in der Regel nicht vorliegt, sollte ein Berechnungsverfahren zur Anwendung kommen, welches es ermöglicht, die Regressionskoeffizienten so zu ermitteln, dass die Beobachtungswerte der Vergangenheit möglichst gering von der Regressionsgeraden entfernt sind. Die gebräuchlichste Methode stellt hierbei die „Methode der kleinsten Quadrate“ dar.<sup>1</sup> Hierbei werden das absolute Glied  $a$  und der Regressionskoeffizient  $b$  dergestalt bestimmt, dass die Summe der quadrierten Abstände der einzelnen Beobachtungspunkte von der Regressionsgeraden minimiert ist.<sup>2</sup> Die Zielfunktion ergibt sich somit wie folgt:<sup>3</sup>

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2 \Rightarrow \min!$$

Durch die Vornahme der partiellen Ableitungen nach  $a$  und  $b$  sowie deren Nullsetzung ergeben sich die beiden sog. Normalgleichungen, anhand derer sich die beiden Parameter  $a$  und  $b$  wie folgt bestimmen lassen:<sup>4</sup>

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n},$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}.$$

Wurden die beiden Parameter  $a$  und  $b$  ermittelt, liegt dem Prüfer mit der Regressionsfunktion nun die Beschreibung einer stochastischen Abhängigkeit vor, die das Verhalten der zu prüfenden Größe in Abhängigkeit von der beeinflussenden Größe in dem betrachteten Zeitraum wiedergibt.<sup>5</sup> Der Prüfer ist anschließend in der Lage, für den im zu

---

<sup>1</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 58.

<sup>2</sup> Vgl. BAMBERG/BAUR (2002), S. 42-43; BLOCHER/WILLINGHAM (1985), S. 181.

<sup>3</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 58; PINNEKAMP/SIEGMANN (2001), S. 139.

<sup>4</sup> Vgl. BAILEY (1981), S. 210; BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 59; SCHLITGEN (2003), S. 418; PINNEKAMP/SIEGMANN (2001), S. 139.

<sup>5</sup> Vgl. DEAKIN/GRANOF (1976), S. 30.

prüfenden Geschäftsjahr vorliegenden Wert der unabhängigen Variable X einen Erwartungswert für die zu prüfende Größe Y zu ermitteln.

Eine Erweiterung der bisher geschilderten Vorgehensweise ist die multiple bzw. Mehrfachregressionsanalyse, bei der im Gegensatz zur Einfachregression nicht nur eine beeinflussende Variable sondern zwei oder mehrere Variablen zur Ermittlung eines Erwartungswerts berücksichtigt werden.<sup>1</sup> Bezogen auf das vorangegangene Beispiel könnte der ausgewiesene Bestand an Forderungen aLuL nicht nur von den Umsatzerlösen der betrachteten Periode abhängig sein, sondern auch von diversen anderen Faktoren, wie etwa der gesamtwirtschaftlichen Lage oder den vereinbarten Zahlungsbedingungen. Durch die zusätzliche Berücksichtigung dieser Faktoren kann der Zuverlässigkeit des berechneten Erwartungswerts im Gegensatz zur Einfachregression erhöht werden.<sup>2</sup> Die entsprechende Formel für eine derartige Mehrfachregression gestaltet sich wie folgt.<sup>3</sup>

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_jX_j + \dots + b_kX_k + e$$

mit

Y = abhängige Variable (hier die zum Geschäftsjahresende ausgewiesenen Forderungen)

$X_1, X_2, \dots, X_k$  = erklärende Variable (hier die Umsatzerlöse des betrachteten Geschäftsjahrs)

a = absolutes Glied

$b_1, b_2, \dots, b_k$  = Regressionskoeffizienten

e = Störvariable

Die Ermittlung des absoluten Glieds a und der Regressionskoeffizienten erfolgt bei der Mehrfachregression analog zur Einfachregression durch die Methode der kleinsten Quadrate.<sup>4</sup>

Die Schwierigkeit bei der Durchführung der Mehrfachregression ist häufig darin begründet, dass zwei bzw. mehrere Einflussfaktoren identifiziert werden müssen, die gegenseitig unabhängig voneinander sind. Dieses Problem der gegenseitigen Beeinflus-

---

<sup>1</sup> Vgl. MCKEE (1989), S. 130-131.

<sup>2</sup> Vgl. KINNEY/UECKER (1986), S. 10; GÄRTNER (1994), S. 147.

<sup>3</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 69; BAILEY (1981), S. 230.

<sup>4</sup> Vgl. GÄRTNER (1994), S. 149.

sung unabhängiger Variablen wird in der Literatur auch als Multikolinearität bezeichnet.<sup>1</sup> In diesem Fall werden die Ergebnisse der Mehrfachregression verfälscht, d.h. der Prüfer erhält einen Erwartungswert, dessen Aussagekraft erheblich beeinträchtigt ist.<sup>2</sup> Wiederum auf das vorangegangene Beispiel bezogen, läge Multikolinearität etwa dann vor, wenn der Prüfer einen Erwartungswert für den Forderungsbestand entwickeln möchte und dabei auf die Umsatzerlöse sowie die Auftragseingänge als beeinflussende Variablen zurückgreift, da davon auszugehen ist, dass die Umsätze auch von der Höhe der Auftragseingänge beeinflusst werden.<sup>3</sup>

#### **2.2.4.2.3 Vorteile und Nachteile der Regressionsanalyse**

Ein wesentlicher Vorteil der Regressionsanalyse liegt neben der strukturierten und somit objektiven Vorgehensweise insbesondere darin, dass die Sicherheit und Genauigkeit des Erwartungswerts prinzipiell exakt quantifiziert werden kann, so dass der Prüfer auf einen eindeutigen und objektiven Maßstab für die Beweiskraft des Prüfungsnachweises zurückgreifen kann.<sup>4</sup>

Als geeignete Parameter zur Einschätzung der Sicherheit (bzw. Zuverlässigkeit) eines Erwartungswerts bieten sich das Bestimmtheitsmaß bzw. der Korrelationskoeffizient<sup>5</sup> an.<sup>6</sup> Das Bestimmtheitsmaß kann dabei als Messgröße für die Erklärungskraft der ermittelten Regressionsfunktion herangezogen werden.<sup>7</sup> Der Wert dieses Parameters kann zwischen null und eins liegen und gibt dabei an, wie viel Prozent der Abweichungen in den Daten durch die Regressionsfunktion erklärt werden.<sup>8</sup> Als cut off-Wert für die Entscheidung, ob eine Regressionsanalyse und somit auch der hieraus abgeleitete Erwartungswert als zuverlässig anzusehen ist, bildet die 70%-Grenze. Regressionsanalysen, die ein Bestimmtheitsmaß von weniger als 0,7 aufweisen, sollten daher vom Prüfer verworfen werden.<sup>9</sup>

---

<sup>1</sup> Vgl. NEWBOLD/BOS (1990), S. 77; MCKEE (1989), S. 138; ARRINGTON/HILLISON/ICERMAN (1983), S. 173.

<sup>2</sup> Vgl. BACKHAUS/ERICHSON/PLINKE/WEIBER (2006), S. 89-90.

<sup>3</sup> Vgl. SPERL (1978), S. 228.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 180; MARTEN/QUICK/RUHNKE (2007), S. 305-306.

<sup>5</sup> Da das Bestimmtheitsmaß lediglich das Quadrat des Korrelationskoeffizienten ist und die Aussagen der beiden Parameter generell in die gleiche Richtung gehen, wird an dieser Stelle ausschließlich auf das Bestimmtheitsmaß eingegangen.

<sup>6</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 183.

<sup>7</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 183.

<sup>8</sup> Vgl. MCKEE (1989), S. 128.

<sup>9</sup> Vgl. MCKEE (1989), S. 129.

Ein geeigneter Parameter zur Einschätzung der Genauigkeit eines anhand der Regressionsanalyse abgeleiteten Erwartungswerts ist die *Standardabweichung*.<sup>1</sup> Die Standardabweichung ermöglicht dem Prüfer, eine Bandbreite um den entwickelten Erwartungswert zu bestimmen, worin der „wahre“ Wert der zu prüfenden Größe mit hoher Wahrscheinlichkeit liegt.<sup>2</sup> Somit steht dem Prüfer mit der Standardabweichung ein Mittel zur Verfügung, welches ihn bei der Beurteilung unterstützt, ob eine festgestellte Abweichung zwischen Erwartungswert und ausgewiesenem Buchwert als wesentlich aufzufassen ist oder nicht.<sup>3</sup> Im Kontext der Plausibilitätsbeurteilung von Jahresabschlussgrößen erscheint eine Festlegung der Toleranzgrenzen auf das zweifache der Standardabweichung ober- und unterhalb des Erwartungswerts als angemessen.<sup>4</sup> Der Prüfer könnte mit 95%-iger Sicherheit davon ausgehen, dass alle Buchwerte, die innerhalb dieser Bandbreite liegen, frei von wesentlichen Fehlern sind. Sollte der ausgewiesene Buchwert außerhalb dieses Intervalls liegen, sollte der Prüfer die Normenkonformität in Frage stellen und weitere auf die Ursache dieser Abweichung gerichtete Prüfungshandlungen durchführen.<sup>5</sup>

Trotz des verstärkten Formalisierungsgrads und der mit der Quantifizierung von stochastischen Zusammenhängen verbundenen erhöhten Objektivität weisen auch Regressionsanalysen einige problematische Aspekte auf, die ihrer Aussagekraft u.U. entgegenstehen können.

Ein wesentlicher Kritikpunkt ist etwa in der *Auswahl der unabhängigen Variablen* zu sehen. Da die Regressionsanalyse ein rein quantitatives Verfahren ist, ermöglicht sie dem Prüfer keine qualitative Beurteilung des Zusammenhangs zwischen den berücksichtigten Variablen.<sup>6</sup> Zwar steht dem Prüfer mit dem Bestimmtheitsgrad bzw. dem Korrelationskoeffizienten eine Größe zur Beurteilung der Verlässlichkeit der jeweiligen Regressionsfunktion zur Verfügung. Allerdings handelt es sich hierbei um eine Messung der rein zahlenmäßigen Verbundenheit der betrachteten Variablen.<sup>7</sup> So besteht durchaus die Möglichkeit, dass der anhand eines Bestimmtheitsgrads bzw. eines Korrelationskoeffizienten gemessene quantitative Zusammenhang durchaus von Bedeutung

---

<sup>1</sup> Vgl. KINNEY/MCDANIEL (1996), S. 53.

<sup>2</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 110. BLOCHER/WILLINGHAM quantifizieren die Wahrscheinlichkeit, dass sich der „wahre“ Wert innerhalb dieser Bandbreite befindet auf 67%. Vgl. BLOCHER/WILLINGHAM (1985), S. 184.

<sup>3</sup> Vgl. SMITH (1989), S. 23.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 184.

<sup>5</sup> Vgl. MCKEE (1989), S. 129.

<sup>6</sup> Vgl. STEWARD (1978), S. 70; DEAKIN/GRANOF (1974), S. 765.

<sup>7</sup> Vgl. MÜLLER (1996), S. 49.

ist, qualitativ betrachtet ein Zusammenhang zwischen den betrachteten Größen jedoch völlig abwegig ist.<sup>1</sup> Der Prüfer wählt die unabhängige(n) Variable(n) daher letztendlich auf Basis sachlogischer Überlegungen aus,<sup>2</sup> was zeigt, dass auch beim Einsatz von Regressionsanalysen nicht auf das persönliche Urteilsvermögen des Prüfers verzichtet werden kann.<sup>3</sup>

Ein weiteres Problem ist in der *Zuverlässigkeit der verwendeten Daten* zu sehen.<sup>4</sup> Einem durch eine Regressionsanalyse ermittelten Erwartungswert kann nur dann eine hohe Aussagekraft bescheinigt werden, wenn die einfließenden Daten zuverlässig sind. Dieses ist tendenziell eher bei geprüften als bei ungeprüften Daten zu vermuten. Da der Prüfer bei Regressionsanalysen allerdings auf die Verfügbarkeit einer breiten Datenbasis angewiesen ist, kommt der Prüfer in der Regel nicht umher, auf Monatswerte zurückzugreifen.<sup>5</sup> Diese weisen jedoch das Problem auf, dass es sich dabei um ungeprüfte Werte handelt, womit die Zuverlässigkeit der hieraus abgeleiteten Erwartungswerte erheblich in Frage gestellt werden kann.<sup>6</sup>

Letztendlich ist eine Erwartungsbildung durch Regressionsanalysen auch deshalb kritisch zu betrachten, da sie von der *Gültigkeit der Regressionsbeziehung* bis in die zu prüfende Periode ausgeht.<sup>7</sup> Ist diese Bedingung jedoch nicht erfüllt, sollte der Prüfer auf den Einsatz von Regressionsanalysen verzichten, da die Beweiskraft ihrer Ergebnisse erheblich gemindert wird.<sup>8</sup>

---

<sup>1</sup> So könnte beispielsweise ein Zusammenhang zwischen den Instandhaltungskosten und dem Zinsertrag auf quantitativer Ebene durchaus gegeben sein. Aus sachlogischen Überlegungen heraus erscheint dieser Zusammenhang jedoch höchst unplausibel. Vgl. GRAHAM (1981), S. 22.

<sup>2</sup> Vgl. GÄRTNER (1994), S. 147.

<sup>3</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 305-306; GÄRTNER (1994), S. 141 sowie 153 m.w.N.

<sup>4</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 190.

<sup>5</sup> Vgl. BLOCHER/WILLINGHAM (1985), S. 190.

<sup>6</sup> Vgl. BAILEY (1981), S. 209.

<sup>7</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 305-306.

<sup>8</sup> Vgl. KINNEY/BAILEY (1976), S. 397.

## 2.3 Ablaufmodell bei der Durchführung analytischer Prüfungshandlungen

### 2.3.1 Die Zerlegung der prüferischen Vorgehensweise in einzelne Prüfungsschritte

Wie bereits in Kap. 2.1.1 skizziert, handelt es sich bei analytischen Prüfungshandlungen um *Soll-Ist-Vergleiche*, bei denen die ungeprüften Buchwerte des Mandanten (sog. *Ist-Werte*) den vom Prüfer entwickelten Erwartungswerten (sog. *Soll-Werte*) gegenübergestellt und dadurch auf ihre Plausibilität beurteilt werden.<sup>1</sup> Stellt der Prüfer keine signifikante Differenz zwischen den beiden Werten fest, so kann er dies als Prüfungsnachweis für die Nomenkonformität des zugrunde liegenden Jahresabschlusswerts heranziehen. Für diese Situation könnte sich der Prüfer etwa für die Reduzierung oder den Wegfall weiterer Prüfungshandlungen *entscheiden*.<sup>2</sup> Weichen die Werte hingegen wesentlich voneinander ab, so könnte dies auf einen Fehler in den Jahresabschlusssdaten hindeuten. In diesem Fall ist der Prüfer dazu verpflichtet, weitere *Nachforschungen* durchzuführen, um den tatsächlichen Grund für die festgestellte Abweichung zu identifizieren.<sup>3</sup> Hat der Prüfer genügend Prüfungsnachweise eingeholt, die den Grund für die beobachtete Abweichung dokumentieren, obliegt es ihm, eine *abschließende Prüfungsentscheidung* zu treffen, welche im Wesentlichen davon abhängen wird, ob es sich bei dem identifizierten Abweichungsgrund um einen Fehler im Datenmaterial oder um einen „natürlichen“ Grund (sog. Nicht-Fehler), wie etwa Änderungen in der Unternehmensumwelt oder der internen Unternehmensstruktur bzw. -organisation,<sup>4</sup> handelt.

Diese Ausführungen verdeutlichen, dass es sich bei analytischen Prüfungshandlungen nicht um eine einzelne Aktivität des Prüfers handelt, sondern vielmehr um eine Kette mehrerer aneinander gereihter Aktivitäten, die in ihrer Gesamtheit als ein Ablaufmodell einer analytischen Prüfungshandlung verstanden werden können.<sup>5</sup> Im weiteren Verlauf der Untersuchung werden diese Aktivitäten als einzelne *Prüfungsschritte* bezeichnet. Als wesentliche Prüfungsschritte lassen sich dabei die *Erwartungsbildung*, die *Festlegung von Toleranzgrenzen*, der *Soll-Ist-Vergleich*, die anschließende *Beurteilung*, wel-

---

<sup>1</sup> Dabei zielt der Soll-Ist-Vergleich nicht darauf ab, eine exakte Gleichheit zwischen den beiden Werten festzustellen, als vielmehr die sachlogische Übereinstimmung (Plausibilität) zu untersuchen. Vgl. IDW (2006), R 371.

<sup>2</sup> Vgl. hierzu ausführlich Kap. 2.1.2.2.1.3.

<sup>3</sup> Vgl. IDW PS 312.26 sowie ISA 520.17 sowie KROMMES (2006), S. 214.

<sup>4</sup> Siehe hierzu insbesondere Kap. 2.3.2.1.2.

<sup>5</sup> So auch BIGGS/MOCK/QUICK (2000), S. 171-173.

che ihrerseits weitere *Nachforschungen* bedingt, und die abschließende *Prüfungsentscheidung* herausstellen. Demnach kann in einer ersten Annäherung von folgender Ablaufstruktur bei der Durchführung von analytischen Prüfungshandlungen ausgegangen werden (siehe Abbildung 3), die es im Folgenden genauer zu spezifizieren gilt.

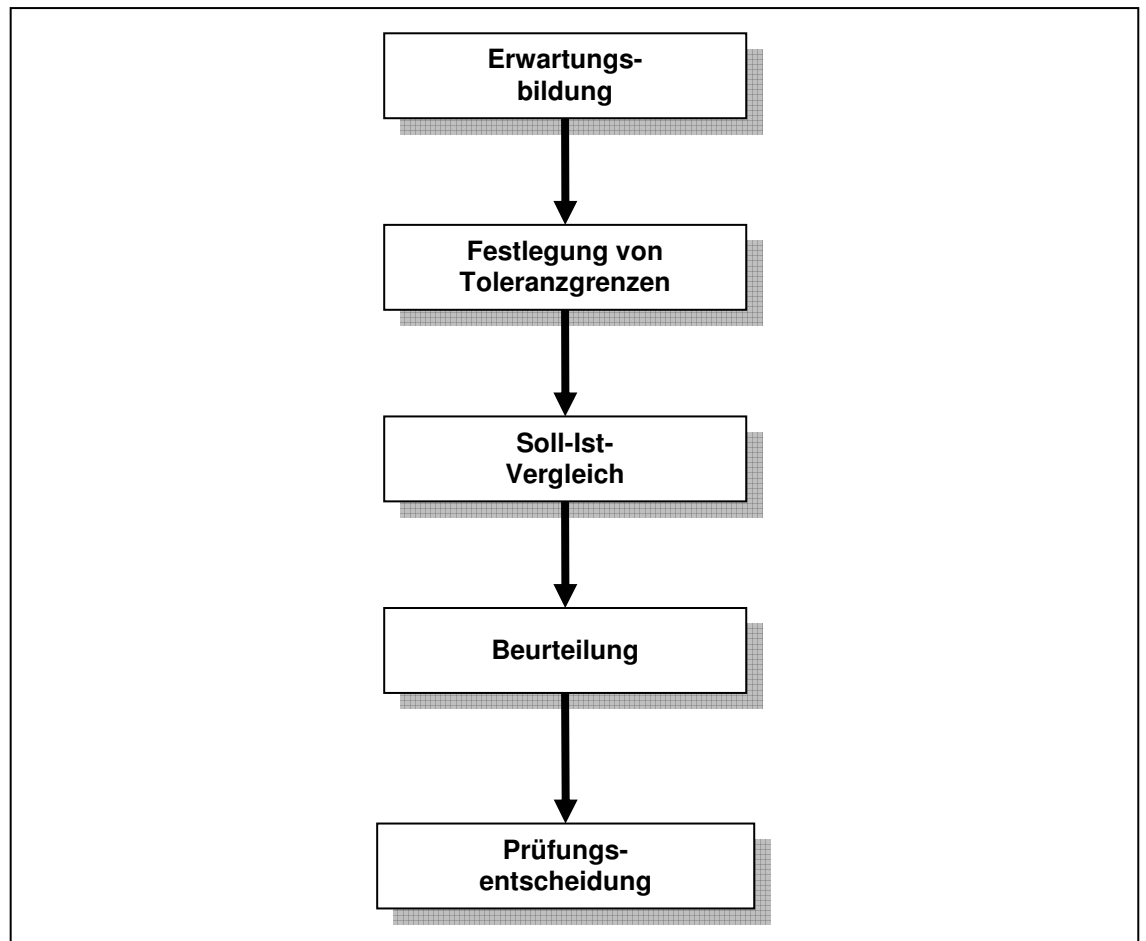


Abbildung 3: Ablaufmodell zur Durchführung von analytischen Prüfungshandlungen

## **2.3.2 Beschreibung der einzelnen Prüfungsschritte bei der Durchführung von analytischen Prüfungshandlungen**

### **2.3.2.1 Bildung eines Erwartungswerts**

#### **2.3.2.1.1 Einflussfaktoren der Qualität des Erwartungswerts**

Die Bildung eines adäquaten Erwartungswerts ist von nicht zu unterschätzender Bedeutung, da sich die Qualität des Erwartungswerts maßgeblich auf den Erfolg der analyti-



schen Prüfungshandlung auswirkt.<sup>1</sup> Dabei wird die Qualität des Erwartungswerts im Wesentlichen von den beiden folgenden Aspekten beeinflusst:

1. *Das Verständnis des Prüfers für die Zusammenhänge in den jahresabschlussrelevanten Daten*
2. *Der Genauigkeit des entwickelten Erwartungswerts.*<sup>2</sup>

Im Folgenden werden diese beiden Punkte genauer spezifiziert.

#### **2.3.2.1.2      Verständnis des Prüfers für die Zusammenhänge in den jahresabschlussrelevanten Daten**

Da es sich bei Jahresabschlussdaten regelmäßig um ein Netzwerk von miteinander in Verbindung stehenden Größen handelt, sollte der Prüfer diese nicht isoliert voneinander, sondern immer mindestens zwei miteinander verknüpfte Werte gemeinsam betrachten. Eine wesentliche Ursache für derartige Zusammenhänge liegt in der doppelten Buchführung.<sup>3</sup> Möchte der Prüfer einen zuverlässigen Erwartungswert bilden, ist es daher unabdingbar, dass er über ein umfassendes Verständnis über die im Datenmaterial befindlichen Zusammenhänge verfügt.<sup>4</sup> In der Literatur wird dieses Wissen auch häufig als rechnungslegungsspezifisches Wissen (accounting knowledge) bezeichnet.<sup>5</sup>

Die vom Prüfer zur Entwicklung eines Erwartungswerts herangezogenen Zusammenhänge im Datenmaterial haben dabei insbesondere die Anforderungen *der Richtigkeit, Relevanz, Beständigkeit und der Beobachtungshäufigkeit* zu berücksichtigen.<sup>6</sup> Dabei spielt es eine besondere Rolle, dass der Prüfer neben rechnungslegungsspezifischen Kenntnissen auch über ein ausreichendes Maß an unternehmens- und umweltspezifischem Wissen (sog. *business knowledge*<sup>7</sup>) verfügt.<sup>8</sup> Im Vordergrund stehen hierbei Kenntnisse über die Unternehmensstruktur und -organisation (sog. *interne Faktoren*)

---

<sup>1</sup> Vgl. IDW (2006), R 372.

<sup>2</sup> Vgl. IDW (2006), R 372.

<sup>3</sup> Vgl. SMITH (1989), S. 17.

<sup>4</sup> Vgl. PENNY (2003), S. 92.

<sup>5</sup> Vgl. etwa BLOCHER (1983), S. 26 oder MCKEE (1989), S. 21.

<sup>6</sup> Vgl. IDW (2006), R 373 sowie SMITH (1989), S. 18.

<sup>7</sup> Vgl. etwa BLOCHER (1983), S. 26 sowie MCKEE (1989), S. 21.

<sup>8</sup> Vgl. KINNEY/MCDANIEL (1996), S. 52.

und das rechtliche und wirtschaftliche Umfeld (sog. *externe Faktoren*) des zu prüfenden Unternehmens sowie diesbezügliche Änderungen im zu prüfenden Geschäftsjahr.<sup>1</sup>

So wäre beispielsweise bei der Entwicklung eines Erwartungswerts für den Provisionsaufwand anhand der Umsatzerlöse die *Richtigkeit des Zusammenhangs* vehement in Frage gestellt, wenn für die Zahlung der Provision nicht nur die Umsätze berücksichtigt würden, sondern auch der Zahlungseingang bei den entsprechenden Forderungen.<sup>2</sup>

Sollten bei einem Erwartungswert für den Personalaufwand lediglich die Tariflohnsteigerungen berücksichtigt werden, so wäre die *Relevanz des Zusammenhangs* in hohem Maße unzureichend, falls der Mandant zu einem nicht unwesentlichen Teil auf außertarifliche Beschäftigungsverhältnisse und gewerbliche Mitarbeiter zurückgreift.<sup>3</sup>

Die *Beständigkeit eines Zusammenhangs* wäre etwa dann entschieden anzuzweifeln, wenn sich in der zu prüfenden Periode interne oder externe Änderungen ergeben haben, die sich maßgeblich auf den Ursache-Wirkungszusammenhang auswirken.<sup>4</sup> So wäre beispielsweise denkbar, dass bei der Bildung eines Erwartungswerts für den Personalaufwand die durch Umstrukturierungen bedingten Änderungen in der Mitarbeiterstruktur zu berücksichtigen sind.<sup>5</sup>

Schließlich wäre die Anforderung der *Beobachtungshäufigkeit eines Zusammenhangs* nur dann erfüllt, wenn der für die Entwicklung eines Erwartungswerts zugrunde gelegte Zusammenhang in der Vergangenheit wiederholt nachgewiesen werden konnte.<sup>6</sup> Sollte sich etwa eine bestimmte Beziehung zwischen zwei Jahresabschlussgrößen nur einmal in den letzten drei Jahren wieder gefunden haben, spiegelt dies wohl keine ausreichende Basis für eine zuverlässige Erwartungsbildung wider.

### **2.3.2.1.3 Genauigkeit des entwickelten Erwartungswerts**

Die Genauigkeit des Erwartungswerts stellt ein Maß für die Verlässlichkeit der analytischen Prüfungshandlung insgesamt dar und beeinflusst maßgeblich die Prüfungssicher-

---

<sup>1</sup> Vgl. PENNY (2003), S. 92. Einen kurzen Überblick über mögliche interne und externe Faktoren gibt PENNY (2003), S. 93. Für eine ausführliche Darstellung siehe hingegen MARTEN/QUICK/RUHNKE (2007), S. 256-269.

<sup>2</sup> Vgl. IDW (2006), R 373.

<sup>3</sup> Vgl. IDW (2006), R 373.

<sup>4</sup> Vgl. SMITH (1989), S. 18.

<sup>5</sup> Vgl. IDW (2006), R 373.

<sup>6</sup> Vgl. SMITH (1989), S. 18 sowie IDW (2006), R 373.

heit, die durch die jeweilige analytische Prüfungshandlung gewährleistet wird.<sup>1</sup> Dabei hängt der Grad der Genauigkeit eines Erwartungswerts im Wesentlichen von der *Präzision des eingesetzten Verfahrens* sowie der *Zuverlässigkeit* und dem *Aggregationsgrad des verwendeten Datenmaterials* ab.<sup>2</sup>

Bei stark vereinfachten Verfahren zur Ermittlung eines Erwartungswerts, wie etwa einem einfachen Vorjahresvergleich, besteht die Gefahr, dass die Prüfungsnachweise eine lediglich geringe Aussagekraft aufbringen, wohingegen kosten- und zeitintensivere Verfahren, wie etwa modellbasierte Verfahren mit mehreren unabhängigen Variablen, zu genaueren Erwartungswerten und somit zu einer höheren Prüfungssicherheit beitragen.<sup>3</sup>

Die Verwendung von unzuverlässigem Datenmaterial kann die Aussagekraft der Prüfungsnachweise ebenfalls erheblich beeinträchtigen, da sich die hieraus abgeleiteten Erwartungswerte kaum als brauchbare Vergleichsobjekte zur Beurteilung der ungeprüften Jahresabschlusswerte heranziehen lassen.<sup>4</sup> Vergleichbar verhält es sich mit dem Aggregationsgrad des Datenmaterials.<sup>5</sup> Erwartungswerte, die auf Basis von hoch aggregierten Daten generiert wurden, etwa auf der Ebene des Gesamtunternehmens für das komplette Geschäftsjahr, weisen in der Regel eine geringere Genauigkeit auf als Erwartungswerte, die auf der Grundlage von detaillierten Daten, etwa Monats- bzw. Quartalswerte für einzelne Produktgruppen, gebildet werden.<sup>6</sup>

### **2.3.2.2 Festlegung von Toleranzgrenzen**

#### **2.3.2.2.1 Zweck und Zeitpunkt der Festlegung von Toleranzgrenzen**

Nachdem der Prüfer einen Erwartungswert für den zu prüfenden Sachverhalt ermittelt hat, sollte er eine Bandbreite tolerierbarer Abweichungsbeträge um den entwickelten Erwartungswert bestimmen.<sup>1</sup> Die Festlegung dieser sog. Toleranzgrenzen stellt einen kritischen Aspekt der gesamten Prüfungshandlung dar, da sie die Basis für die Beurteilung darüber bildet, ob es sich bei einer möglicherweise identifizierten Abweichung um

---

<sup>1</sup> Vgl. IDW (2006), R 374. Zur Verlässlichkeit der Ergebnisse von analytischen Prüfungshandlungen sei auf die Ausführungen zu Kap. 2.1.2.3.1 verwiesen.

<sup>2</sup> Vgl. MÜLLER (1996), S. 55 sowie IDW (2006), R 375.

<sup>3</sup> Vgl. IDW (2006), R 374 sowie KINNEY/MCDANIEL (1996), S. 53. Eine ausführliche Diskussion von Verfahren analytischer Prüfungshandlungen findet sich unter Kap. 2.2.

<sup>4</sup> Vgl. GÄRTNER (1994), S. 40. Zur Zuverlässigkeit der in die analytische Prüfungshandlung einfließenden Informationen wird auf Kap. 2.1.2.3.1.4 verwiesen.

<sup>5</sup> Vgl. im Folgenden CALLAHAN/JAENICKE/NEEBES (1988), S. 60 sowie IDW (2006), R 375.

<sup>6</sup> Eine ausführliche Diskussion über die Zerlegbarkeit von Daten findet sich in Kap. 2.1.2.3.1.2.

einen signifikanten Betrag handelt, der weitergehende Untersuchungen des zugrunde liegenden Sachverhalts erfordert, oder ob ein tolerierbarer Abweichungsbetrag vorliegt, der als Indiz für die Normenkonformität des zugrunde liegenden Sachverhalts aufgefasst werden kann.<sup>2</sup> Um hierbei einen möglichst hohen Grad an Objektivität zu erzielen, kommt es entscheidend darauf an, dass die Festlegung der Toleranzgrenzen gleichzeitig mit oder unmittelbar nach der Erwartungsbildung erfolgt.<sup>3</sup> Keinesfalls sollte der Prüfer die Toleranzgrenzen erst während des eigentlichen Soll-Ist-Vergleichs bestimmen, da so die Gefahr besteht, dass die Kenntnis eines festgestellten Abweichungsbetrags einen Einfluss auf die Festsetzung der Toleranzgrenzen haben könnte.<sup>4</sup>

#### **2.3.2.2.2 Kriterien zur Festlegung von Toleranzgrenzen**

Als wichtige Kriterien, auf die der Prüfer zur Festsetzung von Toleranzgrenzen zurückgreifen kann, kommen die erforderliche *Prüfungssicherheit*, die *Genauigkeit des entwickelten Erwartungswerts* sowie die mit dem jeweiligen Prüfungsauftrag verbundene *Wesentlichkeitsbeurteilung* in Betracht.<sup>5</sup>

Soll die angewandte analytische Prüfungshandlung etwa zu einer geringen oder moderaten Prüfungssicherheit beitragen, kann der maximal zulässige Abweichungsbetrag größer ausfallen als bei Prüfungshandlungen, die zu einer hohen Prüfungssicherheit führen sollen.<sup>6</sup>

Die Genauigkeit des Erwartungswerts weist einen logischen Zusammenhang zu den tolerierbaren Abweichungsbeträgen auf. Denn je höher die Genauigkeit eines Erwartungswerts ist, desto enger können die Toleranzgrenzen gesetzt werden und desto höher ist die Wahrscheinlichkeit, dass signifikante Abweichungen auf wesentliche Fehler im Datenmaterial zurückzuführen sind.<sup>1</sup> Verwendet der Prüfer zur Erwartungsbildung etwa relativ hoch aggregierte Daten, sollte dies zu einer großen Bandbreite tolerierbarer Abweichungswerte führen. Allerdings resultieren hieraus lediglich wenige aussagekräftige Prüfungsnachweise. Daten mit einem relativ geringen Aggregationsgrad tragen hingen-

---

<sup>1</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 297-298; KAYADELEN (2006), S. 27.

<sup>2</sup> Vgl. MÜLLER (1996), S. 56; GÄRTNER (1994), S. 43.

<sup>3</sup> Vgl. MCKEE (1989), S. 33.

<sup>4</sup> Vgl. hierzu MCKEE (1989), S. 32-33. Zum möglichen Einfluss der Kenntnis eines Ist-Werts auf die Festlegung von Toleranzgrenzen siehe insbesondere auch die Forschungsergebnisse in Kap. 3.5.2.2.2.

<sup>5</sup> Vgl. ISA 520.12f. sowie SMITH (1989), S. 24-25.

<sup>6</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 60. Eine vergleichbare Aussage findet sich auch unter ISA 520.12f.

gen zu engeren Bandbreiten und zu einer höheren Prüfungssicherheit bei.<sup>2</sup> Dieser innere Zusammenhang zwischen Erwartungswert und Toleranzgrenzen kann auch als ein weiteres Argument für die möglichst zeitgleiche Bildung dieser Werte herangezogen werden.<sup>3</sup>

Inwiefern der Prüfer die jeweilige Wesentlichkeitsbeurteilung einfließen lassen kann, hängt dabei maßgeblich von dem angewandten Verfahren der analytischen Prüfungshandlung ab. Verwendet der Prüfer etwa das Verfahren der Regressionsanalyse, werden ihm mit der errechneten Standardabweichung bereits objektive Toleranzgrenzen bereitgestellt, die aufgrund ihrer mathematisch-statistischen Berechnung durch ein hohes Maß an Objektivität gekennzeichnet sind.<sup>4</sup> Problematischer stellt es sich allerdings dar, wenn der Prüfer nicht auf diese häufig mit einer hohen Zeit- und Kostenintensität verbundenen komplexen Verfahren, sondern auf einfachere Verfahren zurückgreift, die keine mathematisch-statistische Maßgröße für die Festlegung von Toleranzgrenzen zur Verfügung stellen. Hier ist der Prüfer verstärkt auf „ein ausgeprägtes prüferisches Urteilsvermögen sowie erhebliche betriebliche Sachkenntnis“<sup>5</sup> (sog. *judgmental methods*) angewiesen. Hierüber verfügen in der Regel allerdings ausschließlich Prüfer, die ein erhöhtes Maß an Berufserfahrung aufweisen.<sup>6</sup>

Aufgrund der mit der Anwendung von *judgmental methods* verbundenen hohen Anforderungen an die Person des Abschlussprüfers, kommen häufig auch einfache Entscheidungsregeln, wie etwa die sog. 5%- bzw. 10%-Regel, zum Einsatz.<sup>7</sup> Demnach dürfen die festgestellten Abweichungen nicht mehr als 5% bzw. 10% vom ermittelten Erwartungswert betragen, um von der Normenkonformität des zugrunde liegenden Prüfungssachverhalts ausgehen zu können. Eine weitere Möglichkeit liegt darin, dass die festgestellte Differenz nicht größer als ein bestimmter Prozentsatz von dem für den jeweiligen Prüfungsauftrag geltenden Wesentlichkeitswert sein darf.<sup>8</sup> Da derartige Vorgehensweisen allerdings sehr pauschal sind und keine Berücksichtigung individueller Sachverhalte des Prüfungsauftrags erlauben, führen sie im Gegensatz zu den o.a. mathematisch-

---

<sup>1</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 60.

<sup>2</sup> Vgl. KINNEY/MCDANIEL (1996), S. 53.

<sup>3</sup> Vgl. hierzu Kap. 2.3.2.2.1.

<sup>4</sup> Vgl. GAUNTT/GLEZEN (1997), S. 57; KINNEY/MCDANIEL (1996), S. 53. Vergleiche hierzu auch ausführlich die Diskussion in Kap. 2.2.4.2.3.

<sup>5</sup> Vgl. GÄRTNER (1994), S. 42.

<sup>6</sup> Vgl. KELL/BOYNTON (1992), S. 194.

<sup>7</sup> Vgl. etwa GAUNTT/GLEZEN (1997), S. 58.

<sup>8</sup> Vgl. KINNEY/MCDANIEL (1996), S. 53.

statistischen Berechnungen bzw. judgmental methods zu weniger aussagekräftigen Prüfungsnachweisen.<sup>1</sup>

### 2.3.2.3 Soll-Ist-Vergleich

Im Anschluss an die Ermittlung des Erwartungswerts und der Festlegung von Toleranzgrenzen nimmt der Prüfer einen Soll-Ist-Vergleich vor. Ziel dieses Vergleichs ist es, eine potenzielle Abweichung zwischen den beiden gegenübergestellten Werten zu identifizieren, um daraus Rückschlüsse in Bezug auf die Normenkonformität der Jahresabschlussdaten zu ziehen.<sup>2</sup> Dabei kommt es dem Prüfer nicht darauf an, dass er eine exakte Übereinstimmung feststellen kann, sondern vielmehr darauf, dass der ausgewiesene Ist-Wert plausibel erscheint.<sup>3</sup> Hierbei kommt es entscheidend darauf an, ob sich der zu prüfende Ist-Wert innerhalb der festgelegten Bandbreiten befindet oder nicht.

Stellt der Prüfer im Rahmen des Soll-Ist-Vergleichs keine wesentliche Abweichung fest, so kann grundsätzlich unterstellt werden, dass der zugrunde liegende Sachverhalt wahrscheinlich keinen wesentlichen Fehler enthält. Von einer absoluten Sicherheit kann allerdings nicht ausgegangen werden, denn die Aussagekraft der mittels analytischer Prüfungshandlungen gewonnenen Prüfungsnachweise wird maßgeblich davon beeinflusst, ob die zur Entwicklung des Erwartungswerts herangezogenen Ursache-Wirkungs-Relationen auch tatsächlich den im zu prüfenden Unternehmen gegebenen Verhältnissen entsprechen.<sup>4</sup>

Werden hingegen signifikante Abweichungen festgestellt, ist zu entscheiden, in welcher Art und in welchem Umfang weitere Prüfungshandlungen vorzunehmen sind, um den tatsächlichen Grund für die Abweichung zu identifizieren. Abweichungen können darauf zurückzuführen sein, dass der zu prüfende Sachverhalt einen Fehler enthält.<sup>5</sup> Der Grund für eine Abweichung kann aber auch darin liegen, dass die Erwartungen „schlecht“ sind, weil hinsichtlich bestimmter Annahmen schlichtweg ein Irrtum vorliegt oder bestimmte Veränderungen der internen und externen Faktoren übersehen wurden.<sup>6</sup>

---

<sup>1</sup> Weitere Vorgehensweisen zur Festlegung von Wesentlichkeitsgrenzen finden sich in Kap. 2.2.2.2.3.2 sowie 2.2.4.2.3.

<sup>2</sup> Vgl. MÜLLER (1996), S. 54; GÄRTNER (1994), S. 42.

<sup>3</sup> Vgl. IDW (2006), R 376.

<sup>4</sup> Vgl. WYSOCKI (2003), S. 191-192.

<sup>5</sup> Vgl. MÜLLER (1996), S. 54.

<sup>6</sup> Vgl. KROMMES (2006), S. 214 sowie PENNY (2003), S. 93.

Schließlich kann eine Abweichung auch darin begründet sein, dass der Erwartungswert aufgrund des angewandten Verfahrens bzw. der berücksichtigten Daten zu ungenau ist.<sup>1</sup>

#### **2.3.2.4 Beurteilung**

##### **2.3.2.4.1 Allgemeine Vorgehensweise im Rahmen der Beurteilung einer signifikanten Abweichung**

Stellt der Prüfer im Rahmen des Soll-Ist-Vergleichs eine signifikante Abweichung fest, sollte er zunächst überprüfen, ob ihm bei der Berechnung des Erwartungswerts möglicherweise ein Fehler unterlaufen ist und ob das angewandte Verfahren sowie die getroffenen Annahmen tatsächlich angemessen sind.<sup>2</sup> Kommt er zu dem Entschluss, dass keine Einwände gegen den entwickelten Erwartungswert erhoben werden können, ist er im Folgenden zu „Nachforschungen“ verpflichtet, deren Ziel es ist, die tatsächliche Abweichungsursache zu identifizieren sowie ausreichende und angemessene Prüfungsnachweise zu erlangen.<sup>3</sup> Für die Beurteilung einer identifizierten Abweichung zwischen dem Erwartungswert und den entsprechenden Ist-Werten spielt die Erfahrung des Abschlussprüfers eine bedeutende Rolle.<sup>4</sup>

Die Beurteilung einer identifizierten Abweichung kann durch den Abschlussprüfer in drei wesentlichen Schritten vorgenommen werden.<sup>5</sup> Zunächst sollte der Prüfer Annahmen über die Ursache für die festgestellte Abweichung treffen (sog. *Hypothesengenerierung*). Anschließend sollte der Prüfer Informationen einholen, welche die potenziellen Abweichungsgründe entweder bestätigen oder entkräften (sog. *Informationsbeschaffung*). Schließlich obliegt es dem Prüfer, die in Betracht kommenden Abweichungsursachen vor dem Hintergrund der neu erlangten Informationen zu bewerten und die Wahrscheinlichste zu selektieren (sog. *Hypothesenbewertung*). Im Folgenden werden diese drei Schritte einer genaueren Betrachtung unterzogen.

---

<sup>1</sup> Vgl. IDW (2006), R 376.

<sup>2</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 62.

<sup>3</sup> Vgl. KROMMES (2006), S. 214; BIGGS/MOCK/QUICK (2000), S. 171; MARTEN/QUICK/RUHNKE (2007), S. 298, 300; KAYADELEN (2006), S. 29.

<sup>4</sup> Vgl. hierzu LIBBY/FREDERICK (1990), S. 349-350 sowie QUICK (1999), S. 214. Auf den Einfluss von unterschiedlichen Erfahrungsarten auf die Durchführung von analytischen Prüfungshandlungen wird noch zu einem späteren Zeitpunkt ausführlich eingegangen. Siehe hierzu Kap. 3.5.3.3.1.3.1, Kap. 3.5.3.3.2.1.4 sowie Kap. 3.5.4.4.3.

<sup>5</sup> Vgl. hierzu Schritte 5 bis 7 des Ablaufmodells von BIGGS/MOCK/QUICK (2000), S. 171-172.

#### 2.3.2.4.2 Hypothesengenerierung

Die Generierung einer zutreffenden Hypothese ist von entscheidender Bedeutung für den Erfolg einer analytischen Prüfungshandlung. Hat der Prüfer erst einmal eine zutreffende Annahme über den Grund der identifizierten Abweichung getroffen, kann er weitere Prüfungshandlungen auf den zugrunde liegenden Sachverhalt anwenden, was eine effektive und effiziente Vorgehensweise des Prüfers gewährleistet.<sup>1</sup>

Bei der Generierung von potenziellen Abweichungsursachen stehen dem Prüfer generell zwei Möglichkeiten zur Verfügung.<sup>2</sup> Einerseits kann der Prüfer durch Befragung anderer am Prüfungsprozess beteiligter Personen, etwa das Personal des Mandanten, den potenziellen Grund für die Abweichung erfahren. In der Literatur verwendet man hierfür auch häufig den Begriff der *externen Hypothesengenerierung* (*external hypothesis generation*).<sup>3</sup> Andererseits kann er Ideen für potenzielle Abweichungsursachen auch durch eigenständige Überlegungen entwickeln. Dieses bezeichnet man in der Literatur häufig als *interne Hypothesengenerierung* (*internal hypothesis generation*).<sup>4</sup>

Die Vorteile einer externen Hypothesengenerierung liegen in erster Linie in ihrem möglichen positiven Einfluss auf die Effizienz von analytischen Prüfungshandlungen.<sup>5</sup> Beispielsweise könnte das Management des Mandanten die unerwartete Abweichung im Datenmaterial ebenfalls identifiziert und möglicherweise bereits den Grund eruiert haben. Sollte der Mandant den Grund vielleicht noch nicht endgültig festgestellt haben, könnte er aber zumindest in der Lage sein, geeignete Abweichungsursachen zu entwickeln, da er aufgrund seiner Nähe zum Unternehmen regelmäßig über einen besseren Kenntnisstand hinsichtlich des zugrunde liegenden Sachverhalts verfügt.

Als problematisch ist die externe Hypothesengenerierung allerdings deshalb zu betrachten, weil die erlangten Informationen aufgrund der mangelnden Unabhängigkeit des Mandantenmanagements nicht immer eine zuverlässige Entscheidungsgrundlage darstellen.<sup>6</sup> Der Prüfer muss bei der Mandantenbefragung seine prüferische Skepsis (*professional scepticism*) bewahren und den erhaltenen Erklärungen kritisch gegenüber ste-

---

<sup>1</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 171.

<sup>2</sup> Vgl. im Folgenden KOONCE (1993), S. 63-64 m.w.N.

<sup>3</sup> Synonym verwendet man für externe Hypothesen auch den Begriff der „vererbten“ Hypothese (*inherited hypothesis*) bzw. „vererbten“ Erklärung (*inherited explanation*) Vgl. etwa CHURCH/SCHNEIDER (1993), S. 333 sowie BIERSTAKER/BEDARD/BIGGS (1999), S. 21.

<sup>4</sup> Vgl. KOONCE (1993), S. 63-64.

<sup>5</sup> Vgl. im Folgenden SMITH (1989), S. 30.

<sup>6</sup> Vgl. GÄRTNER (1994), S. 46.



hen.<sup>1</sup> Insbesondere sollte er auch die Integrität und den Sachverstand des Befragten berücksichtigen sowie ein möglicherweise gegebenes Interesse an der Formulierung von unzutreffenden und irreführenden Antworten in Betracht ziehen.<sup>2</sup>

Bei der internen Hypothesengenerierung hingegen verlässt sich der Prüfer ausschließlich auf sein eigenes Urteilsvermögen und seine Kenntnisse über den Mandanten sowie sein rechtliches und wirtschaftliches Umfeld.<sup>3</sup> Ein wesentlicher Vorteil liegt hierbei in dem hohen Maß an Eigenverantwortlichkeit des Prüfers, was ihn insbesondere vor der Beeinflussung durch Dritte, etwa dem Mandantenmanagement, schützt. Anschließend kommt für den Prüfer weiterhin eine Mandantenbefragung in Betracht. Jedoch besteht nach zuvor eigener Hypothesengenerierung die Möglichkeit, die Antworten des Befragten besser zu beurteilen und geeignete Folgefragen zu stellen.<sup>4</sup>

#### **2.3.2.4.3 Informationsbeschaffung**

Nachdem der Prüfer die Hypothesengenerierung abgeschlossen hat, muss er ausreichende und angemessene Informationen erlangen, die seine Hypothesen bestätigen bzw. entkräften. Hierzu kann der Prüfer sowohl auf sein vorhandenes Wissen über den Mandanten und dessen rechtliches und wirtschaftliches Umfeld als auch auf bereits erlangte Prüfungsnachweise zurückgreifen (sog. interne Informationsbeschaffung).<sup>5</sup> Reicht dieses nicht aus, sollte er sich an diejenigen Personen im geprüften Unternehmen wenden, die für den zugrunde liegenden Sachverhalt verantwortlich sind (sog. externe Informationsbeschaffung).<sup>6</sup> Abweichungen in den Umsatzerlösen sollten etwa mit dem Vertriebsleiter und Veränderungen in der Produktion hingegen mit dem Fertigungsleiter diskutiert werden. Kommt für eine wesentliche Erhöhung der Reparatur- und Instandhaltungsaufwendungen etwa eine ungewöhnlich hohe Auslastung der Maschinen in Betracht, könnte sich der Prüfer an die verantwortlichen Maschineneinrichter oder den Leiter der Instandhaltungsabteilung richten.<sup>1</sup>

Bei der externen Informationsbeschaffung nimmt die Art der Fragestellung durch den Prüfer eine bedeutende Rolle ein. Nach Möglichkeit sollten überwiegend offene Fragen

---

<sup>1</sup> Vgl. PENNY (2003), S. 93 sowie DEFLIESE/JAENICKE/O'REILLY/HIRSCH (1990), S. 269.

<sup>2</sup> Vgl. BERNSTEIN (1986), S. 41 /18.

<sup>3</sup> Vgl. MCKEE (1989), S. 36.

<sup>4</sup> Vgl. GAUNTT/GLEZEN (1997), S. 59.

<sup>5</sup> Vgl. SMITH (1989), S. 30.

<sup>6</sup> Vgl. MCKEE (1989), S. 36.

(etwa „*Wie haben sich die diesjährigen Umsätze/Absätze im Vergleich zum Vorjahr entwickelt?*“) gestellt werden, da diese regelmäßig weniger verzerrte Antworten hervorbringen als geschlossene Fragen, die bereits eine antizipierte Antwort in sich tragen können (etwa „*Sind die diesjährigen Umsatzsteigerungen auf die gestiegenen Preise zurückzuführen?*“).<sup>2</sup>

Dabei sollten alle vom Mandantenpersonal erhaltenen Informationen mit angemessener Skepsis betrachtet und durch weitere Prüfungsnachweise verifiziert werden.<sup>3</sup> Sollten etwa die gestiegenen Umsatzerlöse auf Preiserhöhungen zurückgeführt werden, könnte der Prüfer eine Untersuchung der Verkaufspreise auf Rechnungen des alten und des neuen Jahrs durchführen. Macht der Mandant hingegen die verstärkten Marketingaktivitäten für die Umsatzsteigerungen verantwortlich, sollte der Prüfer die Vertriebsaufwendungen inspizieren.

#### **2.3.2.4.4 Hypothesenbewertung**

Im Anschluss an die Informationsbeschaffung müssen die generierten Hypothesen einer neuen Bewertung unterzogen werden, um anschließend diejenige Hypothese herauszufiltern, welche mit der höchsten Wahrscheinlichkeit für die festgestellte Abweichung verantwortlich ist.<sup>4</sup> Problematisch ist an dieser Stelle jedoch, dass Prüfer etwa aufgrund von Unerfahrenheit oder Zeitdruck häufig dazu neigen, voreilig ihre Beurteilung zu beenden, ohne eine vollständige Bewertung der Hypothesen vorgenommen zu haben.<sup>5</sup> Eine Möglichkeit, diesem Problem entgegenzuwirken, besteht darin, dass der Prüfer versucht, die bilanziellen Auswirkungen der in Betracht kommenden Hypothesen zurückzurechnen (sog. *as if entry*<sup>6</sup>), um anschließend zu überprüfen, ob weiterhin wesentliche Abweichungen bestehen.<sup>7</sup> Falls die unterstellten Hypothesen tatsächlich der Wahrheit entsprechen, sollten der „bereinigte“ Buchwert wesentlich weniger vom Erwartungswert abweichen als der tatsächliche Buchwert. Umgekehrt könnte eine inkorrekte Hypothese zu einer höheren Abweichung führen als zuvor.<sup>8</sup>

---

<sup>1</sup> Vgl. SMITH (1989), S. 30.

<sup>2</sup> Vgl. GAUNTT/GLEZEN (1997), S. 59.

<sup>3</sup> Vgl. DEFLIESE/JAENICKE/O'REILLY/HIRSCH (1990), S. 269.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 298, 300.

<sup>5</sup> Vgl. MCKEE (1989), S. 36.

<sup>6</sup> Vgl. MCKEE (1989), S. 36.

<sup>7</sup> Vgl. MCKEE (1989), S. 36.

<sup>8</sup> Vgl. KINNEY (1987), S. 69.

Eine durchaus vergleichbare Vorgehensweise ist die Quantifizierung von Hypothesen.<sup>1</sup> Während *as if entries* rückwärtsgerichtete Berechnungen darstellen, sind Quantifizierungen vorwärtsgerichtet. Sollte der Grund für eine Umsatzerhöhung beispielsweise auf eine Preissteigerung zurückgeführt werden, könnte der Prüfer die Steigerung mit dem Absatzvolumen multiplizieren, um die durch die unterstellte Hypothese verursachte Abweichung zu quantifizieren.<sup>2</sup>

#### **2.3.2.4.5 Prüfungsentscheidung**

Der abschließende Schritt bei der Durchführung von analytischen Prüfungshandlungen ist die Prüfungsentscheidung. Hier hat der Prüfer eine Schlussfolgerung für den zugrunde liegenden Sachverhalt zu ziehen.<sup>3</sup> Wie sich der Prüfer dabei letztendlich entscheidet, ist regelmäßig von dem Ausgang der zuvor vorgestellten Schritte abhängig.

Stellt der Prüfer etwa im Rahmen des Soll-Ist-Vergleichs keine wesentliche Abweichung fest, so kann er drei unterschiedliche Entscheidungen fällen.<sup>4</sup> Erstens könnte der Prüfer entscheiden, dass der erlangte Prüfungsnachweis ausreicht und der zugrunde liegende Sachverhalt als normenkonform akzeptiert wird. Zweitens könnte der Prüfer den erlangten Prüfungsnachweis zum Anlass nehmen, um den Umfang der geplanten Prüfungshandlungen im gleichen Prüffeld zu reduzieren. Und drittens könnte der Prüfer entscheiden, dass der Umfang der geplanten Prüfungshandlungen beibehalten wird. Dies könnte etwa der Fall sein, wenn es sich bereits um einen Minimalplan handelt, so dass keine weiteren Prüfungshandlungen eingeschränkt werden können. Es könnte aber auch sein, dass der Prüfer dem erlangten Prüfungsnachweis eine zu geringe Aussagekraft zuordnet. Ein Grund für eine derartige Prüfungsentscheidung wäre etwa in der geringen Verlässlichkeit der Prüfungshandlungen zu sehen, etwa bei sehr schwachen Verfahren wie dem einfachen Vorjahresvergleich, oder der mangelnden Zuverlässigkeit der eingegangenen Daten.<sup>5</sup>

Stellt der Prüfer im Rahmen des Soll-Ist-Vergleichs hingegen eine wesentliche Abweichung fest, muss er zunächst eine Beurteilung dieser Abweichung vornehmen.<sup>6</sup> In Ab-

---

<sup>1</sup> Vgl. GAUNTT/GLEZEN (1997), S. 59.

<sup>2</sup> Vgl. GAUNTT/GLEZEN (1997), S. 59-60.

<sup>3</sup> Vgl. PENNY (2003), S. 93.

<sup>4</sup> Vgl. im Folgenden BIGGS/MOCK/QUICK (2000), S. 171.

<sup>5</sup> Ähnlich argumentiert MCKEE (1989), S. 37.

<sup>6</sup> Vgl. IDW PS 312.26 bzw. ISA 520.17.

hängigkeit der erlangten Erkenntnisse kann der Prüfer wiederum drei unterschiedliche Prüfungsentscheidungen treffen. Erstens könnte der Prüfer den zugrunde liegenden Sachverhalt als normenkonform annehmen. Diese wäre etwa der Fall, wenn die Untersuchungen ergeben, dass die festgestellte Abweichung durch eine Änderung im Unternehmen oder dessen Umfeld (sog. Nicht-Fehler) verursacht wurde.<sup>1</sup> Zweitens könnte der Prüfer entscheiden, dass Nachbuchungen im Jahresabschluss vorzunehmen sind.<sup>2</sup> Hiervon wäre auszugehen, wenn im Rahmen der Untersuchungen festgestellt wurde, dass die Abweichung auf einen Fehler in der Rechnungslegung zurückzuführen ist.<sup>3</sup> Drittens könnte der Prüfer die Entscheidung treffen, dass weitere Prüfungshandlungen vorzunehmen sind, um die Ursache für die Abweichung festzustellen.<sup>4</sup> Dies wäre dann der Fall, wenn die Untersuchungen zu keinem schlüssigen Ergebnis geführt haben und die Abweichung demnach weiterhin ungeklärt ist.<sup>5</sup> In einer derartigen Situation sollte der Prüfer berücksichtigen, dass es sich bei der Abweichungsursache um einen Fehler handeln kann.<sup>6</sup> Mögliche Prüfungshandlungen, die der Prüfer an dieser Stelle einsetzen kann, sind Detailprüfungen.<sup>7</sup> Denkbar wäre aber auch, weitere analytische Prüfungshandlungen vorzunehmen.<sup>8</sup>

---

<sup>1</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 173 i.V.m. 172 sowie SMITH (1989), S. 31.

<sup>2</sup> Vgl. KROMMES (2006), S. 214.

<sup>3</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 171-173.

<sup>4</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 62.

<sup>5</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 171.

<sup>6</sup> Vgl. CALLAHAN/JAENICKE/NEEBES (1988), S. 62.

<sup>7</sup> Vgl. IDW (2006), R 377.

<sup>8</sup> Vgl. GÄRTNER (1994), S. 46.

## **2.4      Zwischenfazit**

Es konnte gezeigt werden, dass dem Prüfer in allen drei Phasen der Jahresabschlussprüfung eine Vielzahl von unterschiedlichen Verfahren analytischer Prüfungshandlungen zur Verfügung steht. Je nach Beschaffenheit der berücksichtigten Informationen führen die vorgestellten Verfahren zu Prüfungsnachweisen mit unterschiedlicher Aussagekraft. Es liegt letztendlich in der Eigenverantwortung des Abschlussprüfers, welches Verfahren er in welcher Phase der Abschlussprüfung einsetzt. Von wesentlicher Bedeutung ist dabei, dass der Prüfer die Verlässlichkeit der einfließenden Informationen und damit zusammenhängend die Stärken und Schwächen des jeweiligen Verfahrens abschätzen kann. Unter Berücksichtigung dieser beiden Aspekte obliegt es dem Prüfer, die Zuverlässigkeit des ermittelten Erwartungswerts angemessen zu beurteilen.

Allerdings konnte das abschließend skizzierte Ablaufmodell analytischer Prüfungshandlungen zeigen, dass die Ermittlung des Erwartungswerts lediglich der Inhalt des ersten Prüfungsschritts ist. Unter Umständen liefert das Verfahren bereits ein Maß (etwa bei der Regressionsanalyse) oder zumindest erste Anhaltspunkte (etwa bei der Trendanalyse) für die Festsetzung der Toleranzgrenzen. Die Relevanz der Verfahren endet jedoch spätestens nach der Gegenüberstellung des ermittelten Erwartungswerts mit den zu prüfenden Jahresabschlussdaten. Hiernach hat der Prüfer das Ergebnis des Soll-Ist-Vergleichs zu beurteilen, wobei der Prüfer insbesondere bei der Identifikation der Abweichungsursache einer großen Herausforderung gegenübersteht. Die Ausführungen zu den einzelnen Beurteilungsmaßnahmen konnten zwar bereits erste Anhaltspunkte zur prüferischen Vorgehensweise geben, allerdings stellen sie keine ausreichende Basis dar, um Handlungsempfehlungen für die Steigerung der Effektivität und Effizienz von analytischen Prüfungshandlungen auszusprechen. Daher soll im folgenden Kap. auf die weit reichenden Erkenntnisse der empirischen Prüfungsforschung zurückgegriffen werden.

### **3 Analytische Prüfungshandlungen als Gegenstand der empirischen Prüfungsforschung**

#### **3.1 Theoretische und konzeptionelle Vorbemerkungen**

##### **3.1.1 Prüfungstheoretische Grundlagen**

Obwohl die Prüfungslehre zu den ältesten Teildisziplinen der Betriebswirtschaftslehre gehört,<sup>1</sup> herrscht noch weitgehend Uneinigkeit über die Begriffe und Inhalte einer Theorie des Prüfungswesens.<sup>2</sup> Was dabei unter dem Begriff einer *Theorie* im Allgemeinen zu verstehen ist, ist für die Entwicklung und Etablierung einer Prüfungstheorie jedoch von grundlegender Bedeutung. Umso mehr wundert es jedoch, dass dem Theoriebegriff in der prüfungsspezifischen Literatur häufig unterschiedliche Bedeutungen zugeordnet werden. Darüber hinaus wird nicht selten über den Theoriebegriff referiert, ohne diesen Begriff zuvor definiert zu haben.<sup>3, 4</sup>

Für den weiteren Verlauf der vorliegenden Arbeit soll jedoch ein sehr enger Theoriebegriff zugrunde gelegt werden, der auch in der wissenschaftstheoretischen Literatur regelmäßig Verwendung findet.<sup>5</sup> Demnach bezeichnet eine Theorie ein System von miteinander verknüpften *nomologischen Hypothesen*, die zur Erklärung und Voraussage des Verhaltens von Objekten herangezogen werden.<sup>6</sup> Hypothesen stellen Vermutungen über die strukturelle Beschaffenheit der Realität dar. Ist diese Hypothese allgemein und hat sie sich bewährt, so wird sie in den Stand einer nomologischen Hypothese erhoben.<sup>7</sup> Häufig wird anstelle einer nomologischen Hypothese auch der Begriff *Gesetzhypothese* verwandt.<sup>8, 1</sup>

Werden nun diese Ausführungen auf eine Theorie des Prüfungswesens übertragen, so sollte das Ziel aller wissenschaftlichen Bemühungen in diesem Bereich darauf ausgerichtet sein, eine Gesamtheit von Aussagen (nomologische Hypothesen) zu etablieren, die die Abläufe in der Prüfungsrealität einerseits erklären und andererseits alle künfti-

---

<sup>1</sup> Vgl. RICHTER (1997b), S. 251. Vgl. auch RUHNKE (2000), S. 191 m.w.N.

<sup>2</sup> Vgl. RUHNKE (2006), S. 650.

<sup>3</sup> Einen ausführlichen Überblick über die unterschiedliche bzw. nicht definierte Verwendung des Theoriebegriffs in der Literatur zum Prüfungswesen findet sich bei RUHNKE (2000), S. 191-193.

<sup>4</sup> Auf die Mehrdeutigkeit des Theoriebegriffs in der Betriebswirtschaftslehre geht auch SCHANZ (1988), S. 23-24 ein.

<sup>5</sup> Siehe etwa SCHANZ (1988), S. 24; CHMIELEWICZ (1994), S. 87; SEIFFERT (2003), S. 164-168.

<sup>6</sup> Vgl. BORTZ/DÖRING (2002), S. 17; SCHNELL/HILL/ESSER (2005), S. 54; ATTESLANDER (2006), S. 20; SEIFFERT (2003), S. 166-167.

<sup>7</sup> Vgl. SCHANZ (1988), S. 24.

<sup>8</sup> Vgl. stellvertretend SEIFFERT (2003), S. 163.

gen Abläufe in jedweder zeitlich-räumlich unbeschränkten Situation vorhersagen können.

Fragen nach der Vorgehensweise zur Generierung von Aussagen über die Prüfungstheorie beantwortet die Wissenschaftstheorie.<sup>2</sup> Sie ist die „Theorie von der Wissenschaft überhaupt“<sup>3</sup> und wird häufig als eine metawissenschaftliche Disziplin bezeichnet.<sup>4</sup> Sie kann demnach als eine Wissenschaft über die Wissenschaft aufgefasst werden.<sup>5</sup> Als Metawissenschaft liegt ihr Anliegen darin, Vorschläge und Verfahrensregeln zu entwickeln, wie Wissenschaftler zur Erreichung ihrer Forschungsziele vorgehen sollten.<sup>6</sup> Sie ist als Lehre von den Forschungswegen zu verstehen, die darüber informiert, welche Forschungstätigkeiten welche Forschungsergebnisse nach sich ziehen.<sup>7</sup> Darüber hinaus besitzt sie insofern normativen Charakter, als dass Forschungsaktivitäten, die sich nicht an ihre Vorschläge halten, Gefahr laufen, auf eine geringe Akzeptanz in ihrer jeweiligen Disziplin zu stoßen.<sup>8</sup>

Wesentlich für die Wissenschaftstheorie ist, dass sie Auskunft darüber gibt, welche *Forschungsziele* anzustreben sind, und in einem nächsten Schritt erläutert, welche Vorgehensweisen bzw. Programme hierzu geeignet sind. Daher soll zunächst eine ausführliche Darstellung der möglichen Forschungsziele erfolgen, um darauf aufbauend aufzeigen zu können, welche unterschiedlichen *Forschungsprogramme* dem Forscher zur Erreichung seiner Ziele zur Verfügung stehen.

---

<sup>1</sup> Auf das Problem nomologischer Hypothesen wird unter Kap. 3.2.1.2.1.2 noch genauer eingegangen.

<sup>2</sup> In der Literatur werden die Begriffe Wissenschaftstheorie und Wissenschaftslehre häufig synonym verwendet. Vgl. etwa SCHANZ (1988), S. 1.

<sup>3</sup> SEIFFERT (2003), S. 17.

<sup>4</sup> Vgl. etwa CHMIELEWICZ (1994), S. 34.

<sup>5</sup> Vgl. RUHNKE (2000), S. 194.

<sup>6</sup> Vgl. STRÖKER (1987), S. 110.

<sup>7</sup> Vgl. STÄHLIN (1973), S. 1.

<sup>8</sup> Vgl. STRÖKER (1987), S. 110.

### 3.1.2 Forschungsziel<sup>1</sup>

Forschungsziele lassen sich grundsätzlich in vier aufeinander aufbauende Zielarten differenzieren:<sup>2, 3</sup>

- Essentialistisches Forschungsziel (Begriffslehre);
- Theoretisches Forschungsziel;
- Technologisches Forschungsziel;
- Normatives Forschungsziel.

Das *essentialistische Forschungsziel* ist allein auf die Bildung und Präzisierung von Begriffen und Definitionen gerichtet.<sup>4</sup> Eine theoretische Fundierung der Aussagen wird nicht angestrebt, „d.h. es besteht weder die Absicht, die Prüfungsrealität deskriptiv zu erfassen noch sie zu erklären“.<sup>5</sup> Das Erkennen realer Zusammenhänge ist daher weniger von Bedeutung als das Wesen des Erkenntnisobjekts.<sup>6</sup> Die Fortschrittsfähigkeit des es-

---

<sup>1</sup> CHMIELEWICZ (1994), S. 8-18 ordnet die noch zu erläuternden Forschungsziele unterschiedlichen *Forschungskonzeptionen* zu, wobei er unter Forschungskonzeptionen „den Versuch einer Systematik für verschiedene als solche bekannte und eingeführte Forschungsrichtungen, die in der Wirtschaftswissenschaft und in anderen wissenschaftlichen Disziplinen auftreten“ (S. 8), versteht. Dabei ist unverständlich, weshalb er die Begriffe „Forschungsrichtungen“ und „Forschungsstrategien“ als Synonyme für den Begriff „Forschungskonzeptionen“ verwendet, da Erstgenanntes ja gerade in seine Definition von Forschungskonzeptionen einfließt. Der Begriff „Forschungsstrategien“ findet sich indes bei KÖHLER (1978), S. 188 wieder. Nach KÖHLER legen Forschungsstrategien im Wesentlichen die Abfolge verschiedener Arbeitsschritte fest, von denen sich der Forscher eine Erfüllung seiner Forschungsziele erhofft. Eine derart verstandene Forschungsstrategie ist sinngemäß jedoch nicht vergleichbar mit der Definition von Forschungskonzeptionen nach CHMIELEWICZ (1994). Somit ist auch die bei CHMIELEWICZ festgestellte synonyme Verwendung der Begriffe „Forschungskonzeptionen“ und „Forschungsstrategien“ missverständlich. Um Verständnisprobleme zu vermeiden, finden die Begrifflichkeiten „Forschungskonzeptionen“, „Forschungsrichtungen“ und „Forschungsstrategien“ in der vorliegenden Untersuchung keine weitere Berücksichtigung. In diesem Kapitel hält der Autor ausschließlich an den Forschungszielen fest.

<sup>2</sup> Diese Einteilung findet sich bei CHMIELEWICZ (1994), S. 8-15. Eine ähnliche Einteilung findet sich bei KÖHLER (1978), 186-188. KÖHLER verzichtet jedoch auf die Begriffslehre (essentialistisches Forschungsziel). ULRICH (1982), S. 3-4 sowie STÄHLIN (1973), und BEHRENS (1993), Sp. 4768-4769 hingegen differenzieren lediglich ein theoretisches und ein technologisches Forschungsziel.

<sup>3</sup> Nach RICHTER (1997a), S. 17-18 ist auch die sog. Berufskunde den Forschungszielen zuzuordnen, welche sich in erster Linie mit den für den jeweiligen Berufsstand (etwa Prüferberuf) und die jeweiligen Institutionen relevanten rechtlichen und faktischen Rahmenbedingungen beschäftigt. Das Erkenntnisinteresse ist vom Anspruch her überwiegend deskriptiv ausgerichtet. Häufig sind ihre Aussagen normativ und interessenspolitisch geprägt und im Sinne einer Politikberatung bzw. Lobbyismus an den Gesetzgeber adressiert. Da jedoch der überwiegend deskriptive Charakter dieses Forschungsziels im Gegensatz zur aktuellen Forschungspraxis im Prüfungswesen steht, wird dieses Ziel im Folgenden nicht weiter erörtert.

<sup>4</sup> Vgl. CHMIELEWICZ (1994), S. 8.

<sup>5</sup> RICHTER (1997a), S. 16.

<sup>6</sup> Vgl. WÖHE (1959), S. 36; FISCHER-WINKELMANN (1975), S. 18.



sentialistischen Forschungsziels ist gering. Sie hat ihre Daseinsberechtigung in erster Linie in den anfänglichen Stadien einer Wissenschaft, wenn noch begrenztes theoretisches Wissen vorliegt.<sup>1</sup> Für Wissenschaften in einem fortgeschrittenen Stadium ist diese Forschungskonzeption von geringer Relevanz, da sie keinerlei Hilfen bietet, „um die Qualität von Aussagen beurteilen zu können oder zwischen unterschiedlichen Handlungsempfehlungen entscheiden zu können“.<sup>2</sup> Für RUHNKE erscheint es für den Prüfungsbereich durchaus folgerichtig, dass dem essentialistischen Forschungsziel keine größere Bedeutung mehr beigemessen wird.<sup>3</sup>

Das *theoretische Forschungsziel* sieht seine Aufgabe im Gegensatz zum essentialistischen Ansatz nicht nur in der Begriffserklärung. Vielmehr verwendet es in der Begriffslehre entwickelte Begriffe und Definitionen als Bausteine zur Bildung leistungsfähiger Theorien.<sup>4</sup> Demnach steht die Begriffsbildung in der Regel vor der Theoriebildung.<sup>5</sup> Das Erkenntnisinteresse des im Mittelpunkt der wissenschaftstheoretischen Diskussion stehenden theoretischen Forschungsziels ist das Konstatieren von generellen Ursache-Wirkungszusammenhängen in der uns umgebenden Realität.<sup>6</sup> POPPER bezeichnet eine Theorie metaphorisch als ein „Netz, das wir auswerfen, um die Welt einzufangen, - sie zu rationalisieren, zu erklären und zu beherrschen“.<sup>7</sup> Leistungsfähige Theorien verlangen dabei einen hohen Allgemeinheitsgrad und eine hohe Bestätigungsfähigkeit, um Erklärungen und Prognosen als Ziel zu ermöglichen. Konsequenz dieses allgemeinen Theorieverständnisses ist die hohe Kritisierbarkeit von Theorien.<sup>8</sup>

Das *technologische Forschungsziel* rückt wirtschafts- bzw. prüfungsspezifische Sachverhalte und Entscheidungen in den Mittelpunkt der Betrachtung und ist daher durch einen geringeren Allgemeinheitsgrad gekennzeichnet. Im Gegensatz zum theoretischen Forschungsziel werden hier keine theoretischen Ursache-Wirkungszusammenhänge geklärt, als vielmehr „das Ziel/Mittel-Denken, das in der Wirtschaftspraxis als Prozess von Zielsetzungs- und Zielerreichungsentscheidungen auftritt [...], als konstitutives Element auch in die Wissenschaft einbezogen“.<sup>9</sup> Wesensmerkmal des technologischen For-

---

<sup>1</sup> Vgl. SCHANZ (1978), S. 11; FISCHER-WINKELMANN (1981), S. 70; ähnlich argumentiert RUHNKE (2000), S. 197.

<sup>2</sup> RICHTER (1997a), S. 17.

<sup>3</sup> Vgl. RUHNKE (2000), S. 197.

<sup>4</sup> Vgl. CHMIELEWICZ (1994), S. 11.

<sup>5</sup> Vgl. RUHNKE (2000), S. 197.

<sup>6</sup> Vgl. CHMIELEWICZ (1994), S. 11.

<sup>7</sup> POPPER (1994), S. 31.

<sup>8</sup> Vgl. KIESER/KUBICEK (1978), S. 15.

<sup>9</sup> CHMIELEWICZ (1994), S. 169.

schungsziels ist, dass es auf den entwickelten theoretischen Aussagen aufbaut und diese umformt, indem die Wirkungen als Ziele anzustreben und die Ursachen als Mittel herbeizuführen sind.<sup>1</sup> Sie greifen also auf Theorien zurück und übertragen diese pragmatisch auf die Wirtschafts- bzw. Prüfungspraxis.<sup>2</sup> Technologien orientieren sich auf das zielerreichende Gestalten und dienen somit dem Gestaltungsinteresse.<sup>3</sup>

Das *normative Forschungsziel* geht über die bisher aufgeführten Forschungsziele hinaus, indem es eine formulierte technologische Aussage zugrunde legt und anhand eines Werturteils eine klare Stellung zu den verfolgten Zielen, den eingesetzten Mitteln und den möglichen Nebenwirkungen abgibt.<sup>4</sup> Es zeigt also im Gegensatz zur Technologie demnach nicht nur was „machbar“ ist, sondern auch was „sein soll“. Dabei stellen die Begutachtung der Vereinbarkeit verschiedener Ziele sowie die Voraussagbarkeit der faktischen Auswirkungen bestimmter Wertgrundsätze wichtige Auseinandersetzungen mit dem Wertproblem dar.<sup>5</sup> Werturteile charakterisieren sich dadurch, dass (1) der avisierte Sachverhalt in positiver oder negativer Weise ausgezeichnet wird, (2) die Gültigkeit eines normativen Prinzips unterstellt wird, welches ein entsprechendes Verhalten verlangt, und (3) erwartet wird, dass sich die Adressaten der Aussage mit diesem Prinzip identifizieren und daher ein entsprechendes Verhalten annehmen.<sup>6</sup>

Zusammenfassend kann konstatiert werden, dass die vorgestellten Forschungsziele generell zu zwei Oberzielen zusammengefasst werden können. Dabei handelt es sich einerseits um das Erkenntnis- und andererseits um das Gestaltungsziel. Dem *Erkenntnisziel* ist in erster Linie das theoretische Forschungsziel zuzuordnen. Da das Beschreiben der Realität und deren Zusammenhänge eine adäquate Terminologie benötigt, ist die Begriffskunde (essentialistisches Forschungsziel) der theoriegeleiteten Forschung vorgeschaltet und ist somit ebenfalls dem Erkenntnisziel dienlich. Eine Zuordnung zum *Gestaltungsziel* scheint in erster Linie bei der technologiegeleiteten Forschung gerechtfertigt. Die Bereitstellung von Handlungsalternativen zur Erreichung eines angestrebten Zustands ist unmittelbar auf die Beeinflussung der Realität ausgerichtet. Auch dem normativen Forschungsziel ist ein Gestaltungsinteresse immanent. Im Gegensatz zur technologiegeleiteten Forschung umfassen die Handlungsempfehlungen darüber hinaus

---

<sup>1</sup> Vgl. CHMIELEWICZ (1994), S. 11.

<sup>2</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1536.

<sup>3</sup> Vgl. CHMIELEWICZ (1994), S. 11 sowie RUHNKE (2000), S. 193.

<sup>4</sup> Vgl. CHMIELEWICZ (1994), S. 14.

<sup>5</sup> Vgl. KÖHLER (1978), S. 188.

<sup>6</sup> Vgl. ALBERT (1972), S. 148.

noch Werturteile des Aussagenden bezüglich der eingesetzten Mittel und angestrebten Ziele.

Alle vorgestellten Forschungsziele dienen ausschließlich der Etablierung von Aussagen, die in einem nächsten Schritt zu einer Theorie, etwa des Prüfungswesens, zusammengefasst werden können. Sie geben dabei im Allgemeinen Antworten auf die Frage, welche *Art* von Aussagen gewonnen werden können und welche *Kriterien* bei ihrer Formulierung einzuhalten sind. Offen bleibt hingegen die Frage, *wie* bzw. auf welche *Art und Weise* die unterschiedlichen Aussagen gewonnen werden können. Diese forschungsbezogenen Vorgehensweisen werden im Folgenden als *Forschungsprogramme* bezeichnet und einer genaueren Erörterung unterzogen.

### 3.1.3 Forschungsprogramme<sup>1</sup>

Unter Forschungsprogrammen sind um gewisse inhaltliche und/oder methodologische Leitideen organisierte natürliche Beurteilungseinheiten zu verstehen.<sup>2</sup> Dabei weisen Leitideen den Forscher darauf hin, wie dem Forschungsobjekt „zu Leibe zu rücken ist“.<sup>3</sup> Sie lassen sich auch als systemkonstituierende Grundgedanken auffassen, die als Wegweiser herangezogen werden können.<sup>4,5</sup>

Allgemein lassen sich Forschungsprogramme in zwei Kategorien einteilen. Dabei handelt es sich einerseits um *entscheidungslogische* bzw. *formal-analytische* und andererseits um *erfahrungswissenschaftliche* bzw. *empirische* Forschungsprogramme.<sup>6</sup>

Entscheidungslogische Programme sind in erster Linie durch die *deduktive* Methode, die Modellbildung sowie die Verhaltensprämissen des homo oeconomicus charakterisiert.<sup>7</sup> Sie versuchen anhand formaler Modelle, Erkenntnisse über die Realität abzuleiten bzw. das Verhalten einzelner Individuen, Gruppen oder Institutionen vorauszusagen.<sup>1</sup> Im Mittelpunkt stehen dabei überwiegend abstrakte Beschreibungen von Problemstrukturen, wie etwa Handlungs- und Entscheidungsprobleme der am betrachteten Sachver-

---

<sup>1</sup> Der Begriff des Forschungsprogramms findet sich erstmalig bei LAKATOS (1974), S. 89-189.

<sup>2</sup> Vgl. RUHNKE (2006), S. 650.

<sup>3</sup> Vgl. RUHNKE (2000), S. 195.

<sup>4</sup> Vgl. SCHANZ (1988), S. 88.

<sup>5</sup> Forschungsprogramme weisen eine große Ähnlichkeit zu den bei KÖHLER (1978), S. 188 aufgeführten Forschungsstrategien auf, da beide Empfehlungen hinsichtlich des forschersichen Vorgehens aussprechen. Vgl. hierzu auch die Diskussion unter Kap. 3.1.2.

<sup>6</sup> Vgl. RICHTER (1997a), S. 19-20.

<sup>7</sup> Vgl. RICHTER (1997a), S. 19.

halt beteiligten Personen.<sup>2</sup> Formale Modelle können optimale Vorgehensweisen aufzeigen, indem sie Diskrepanzen zwischen rationalem und realem Verhalten aufzeigen.<sup>3</sup> Werden im Rahmen der Modellbildung überwiegend geeignete realitätsnahe Annahmen berücksichtigt, so können formale Modelle zur Lösung praxisrelevanter Probleme beitragen.<sup>4,5</sup>

Im Unterschied zu entscheidungslogischen Forschungsprogrammen sind erfahrungswissenschaftliche Forschungsprogramme durch die *induktive* Methode gekennzeichnet. Durch Beobachtung der Realität sollen frühere sowie gegenwärtige Prüfungsprozesse erfasst und erklärt werden.<sup>6</sup> Die hieraus gewonnenen Erkenntnisse werden zu einem geeigneten System empirisch gehaltvoller Aussagen zusammengeführt. Haben sich diese Aussagen in der Wirklichkeit bewährt, werden diese zu einer Theorie zusammengefasst.<sup>7</sup> Können sie sich hingegen nicht bewähren, so werden sie „vom Sockel der Wissenschaft gestürzt und schlicht als falsche Behauptungen ausgeschieden.“<sup>8</sup>

## **3.2 Weiterführende Überlegungen**

### **3.2.1 Einordnung des (Forschungs-)Objekts „Prüfungswesen“ in die theoretischen und konzeptionellen Grundlagen**

#### **3.2.1.1 Prüfungswesen als realwissenschaftliche und angewandte Wissenschaft zugleich**

Als Teildisziplin der Betriebswirtschaftslehre beschäftigt sich die Prüfungslehre mit *realen* Phänomenen der Erfahrungswelt und kann folglich gleichfalls wie die Betriebs-

---

<sup>1</sup> Ähnlich MARTEN/QUICK/RUHNKE (2007), S. 47.

<sup>2</sup> Vgl. GROCHLA (1978), S. 85.

<sup>3</sup> Vgl. SCHREIBER (2000), S. 15.

<sup>4</sup> In der prüfungsbezogenen Forschungspraxis ist jedoch zu beobachten, dass häufig auch unrealistische Annahmen in ein formales Modell einfließen, etwa wenn bei spieltheoretischen Ansätzen von einer perfekten Prüfungstechnologie ausgegangen wird, deren Einsatz es dem Prüfer ermöglicht, sämtliche Fehler beim Mandanten aufzudecken; dies ist in der Prüfungsrealität mit Blick auf die kriminellen Aktivitäten im Rahmen von Bilanzmanipulationen gerade nicht der Fall. Vgl. RUHNKE (2006), S. 652.

<sup>5</sup> Zum Problem unrealistischer Annahmen sei an dieser Stelle auf ALBERT (1996), S. 451-486 verwiesen.

<sup>6</sup> Vgl. RICHTER (1997a), S. 19. In einem engen Zusammenhang mit der Erfassung steht auch die Voraussage künftiger Abläufe in der Prüfungsrealität. Vgl. RUHNKE (2006), S. 651.

<sup>7</sup> Vgl. RUHNKE (2006), S. 651.

<sup>8</sup> WITTE (1985), S. 179.

wirtschaftslehre als eine *Realwissenschaft* bezeichnet werden.<sup>1</sup> Das Interesse einer realwissenschaftlich orientierten Prüfungslehre liegt in erster Linie darin, die Prozesse in der Prüfungsrealität zu erkennen und zu erfassen.<sup>2</sup> Da das Erkenntnisinteresse der theoriegeleiteten Forschungskonzeption immanent ist,<sup>3</sup> erscheint es folgerichtig, dass eine realwissenschaftlich ausgerichtete Prüfungslehre in erster Linie theoriegeleitete Aussagen konstatieren sollte.<sup>4</sup>

Gleichfalls ist gemeinhin anerkannt, dass die Betriebswirtschaftslehre und mit ihr auch die Prüfungslehre als anwendungsorientierte Wissenschaft zu verstehen ist.<sup>5</sup> Die Aufgabe einer *anwendungsorientierten* Wissenschaft liegt in erster Linie darin, Problemlösungen (etwa Regeln, Modelle oder Verfahren) für praxisbezogenes Handeln zu entwickeln, wobei sie sich dabei auf die Erkenntnisse der theoretischen Wissenschaft sowie die Erfahrungen der Praxis stützt.<sup>6</sup> Demnach muss einer als anwendungsbezogene Wissenschaft verstandenen Prüfungslehre ein ausgeprägtes Gestaltungsinteresse zugestanden werden. Da die Gestaltung der Realität wohlbekannt eine Aufgabe von Technologien ist,<sup>7</sup> erscheint es ebenso gerechtfertigt, dass eine anwendungsorientierte Prüfungslehre technologiegeleitete Aussagen hervorbringen soll.

Demnach erscheint für die Prüfungslehre eine Trennung zwischen einerseits theoretischem und andererseits technologischem Forschungsziel wenig zweckdienlich.<sup>8,9</sup> Diese Auffassung erscheint insbesondere im Hinblick auf die in der Prüfungsliteratur häufig aufgezeigte Nähe des Prüfungsprozesses zur Prüfungslehre gestärkt. VOB etwa ist der Ansicht, dass die praktische Ausübung des Prüferberufs ohne „wissenschaftliches Rüstzeug“ heute kaum möglich sei.<sup>10</sup> WITTE geht einen Schritt weiter und zeigt auf, dass der Prüfungsprozess nicht zu verkennende Parallelen zum wissenschaftlichen Untersu-

---

<sup>1</sup> Hinsichtlich der Betriebswirtschaftslehre soll an dieser Stelle auf SCHANZ (1988), S. 9-16 und CHMIELEWICZ (1994), S. 33-34 verwiesen werden. In Bezug auf die Prüfungslehre vgl. RUHNKE (2000), S. 193.

<sup>2</sup> Vgl. RUHNKE (2006), S. 650.

<sup>3</sup> Siehe hierzu Kap. 3.1.2.

<sup>4</sup> Dass eine Theorie nur dann im wissenschaftlichen Prozess eindeutig kommuniziert werden kann, mithin das essentialistische Forschungsziel der Theoriebildung vorgeschaltet ist, wird an dieser Stelle selbstverständlich vorausgesetzt.

<sup>5</sup> Vgl. LEHNER (1996), S. 73; SCHNEIDER (1995), S. 141-142; SIEBEN (1998), S. 5.

<sup>6</sup> Vgl. ULLRICH/FLURI (1992), S. 200.

<sup>7</sup> Siehe hierzu Kap. 3.1.2.

<sup>8</sup> Ähnlich argumentiert SCHREIBER (2000), S. 27.

<sup>9</sup> RUHNKE (2000), S. 193 konstatiert diesbezüglich, dass eine sachgerechte Gestaltung der Realität das Erkennen derselbigen voraussetzt und dass das Erkenntnisinteresse dem Gestaltungsinteresse somit vorgelagert ist. Dieses steht im Einklang mit der in dieser Arbeit vertretenen Auffassung, dass das theoriegeleitete Forschungsziel i.d.R. die Basis für das technologiegeleitete Forschungsziel ist. Siehe hierzu Kap. 3.1.2.

<sup>10</sup> Vgl. VOB (1926), S. 173-174.

chungsprozess aufweise. „Auch der Praktiker formuliert – ganz ähnlich wie der Wissenschaftler – Hypothesen über die Wirklichkeit. (...) Ohne derartige realitätsbezogene Hypothesen könnte kein Praktiker faktisch handeln.“<sup>1</sup> Zuletzt fasste RICHTER diesen Gedanken auf und bezeichnet die Praxis als „Fallstudie“ der Theorie, in der die generalisierenden theoretischen Aussagen letztendlich konkretisiert werden.<sup>2</sup>

Abschließend bleibt zu erörtern, inwiefern die Prüfungslehre auch einem normativen Forschungsziel gerecht werden kann. Diese Aussagen zeigen im Unterschied zu Technologien nicht nur Handlungsalternativen für die Prüfungspraxis auf, sondern beurteilen darüber hinaus auch die eingesetzten Mittel und/oder die angestrebten Ziele und sprechen nicht selten eine Empfehlung aus.<sup>3</sup> Da sie offensichtlich auf die Gestaltung der Geschehnisse in der Realität abzielt, ist ihr eine besondere Nähe zur technologiegeleiteten Forschungskonzeption zuzusprechen. Insofern wäre das normative ebenso wie das theoriegeleitete Forschungsziel als ein Charakteristikum einer angewandten Prüfungslehre aufzufassen.

In der Prüfungsforschung ist eine derartig ausgerichtete normative Konzeption durchaus regelmäßig anzutreffen. Häufig betrifft sie die Gestaltung und Festsetzung von Prüfungsnormen. Dabei beziehen sich die Empfehlungen gleichwohl auf die Normierung des Prüfungsprozesses (etwa im Rahmen der Durchführung von analytischen Prüfungshandlungen<sup>4</sup>) als auch des Markts für Prüfungsleistungen bzw. des Prüfungsumfelds (etwa zur Einführung einer Rotationspflicht oder zur Trennung von Prüfung und Beratung<sup>5</sup>).<sup>1</sup> Insofern erscheint es durchaus angebracht, dass eine anwendungsorientierte Prüfungslehre auch normative Aussagen formulieren soll.

Diesen Ausführungen folgend ist einer zugleich als realwissenschaftlich als auch als angewandte Wissenschaft verstandenen Prüfungstheorie ein sehr breites Aufgabenspektrum zuzuordnen. Sie soll zunächst die Prüfungsrealität mittels theoriegeleiteter Aussa-

---

<sup>1</sup> WITTE (1985), S. 179.

<sup>2</sup> Vgl. RICHTER (1997a), S. 15.

<sup>3</sup> Siehe hierzu Kap. 3.1.2.

<sup>4</sup> Hierzu gibt es etwa Untersuchungen die auf einen negativen Einfluss der Wortwahl bei der Formulierung von Prüfungsstandards (siehe hierzu ausführlich Kap. 3.5.2.2.1), der Kenntnis des ungeprüften Buchwerts (siehe hierzu Kap. 3.5.2.2.2) oder des Erhalts einer Mandantenerklärung (siehe hierzu etwa Kap. 3.5.2.3, 3.5.3.3.2.1 oder 3.5.5.3.1) hinweisen.

<sup>5</sup> Einen ausführlichen Überblick über empirische Forschungsaktivitäten zur Trennung von Prüfung und Beratung und deren Einfluss auf die Unabhängigkeit des Abschlussprüfers gibt QUICK (2006), S. 48-56. Eine kurze Darstellung der empirischen Forschungsergebnisse zur externen Pflichtrotation und deren Einfluss auf die Unabhängigkeit findet sich bei QUICK (2004), S. 496-499.

gen erfassen, um anschließend Handlungsempfehlungen für die Gestaltung eines möglichst effektiven und effizienten Prüfungsprozesses aussprechen zu können. Darüber hinaus sieht sie ihre Aufgabe auch darin, den Standardsettern bei der Normierung der Abschlussprüfung wichtige Erkenntnisse zu liefern und durch Abgabe eigener Werturteile gestalterisch auf die Normensetzung Einfluss zu nehmen.

Folgt man der Auffassung, dass die Prüfungslehre sowohl Real- als auch eine angewandte Wissenschaft ist, ist in einem nächsten Schritt zu beantworten, wie ein derart anspruchsvoller Aufgabenkatalog adäquat erfüllt werden kann. Es stellt sich also die Frage nach einem geeigneten und leistungsfähigen Forschungsprogramm, welches in der Lage ist, diesen durchaus hohen Ansprüchen gerecht zu werden. In Betracht kommen diesbezüglich in erster Linie formal-analytische bzw. entscheidungslogische Programme oder erfahrungswissenschaftliche bzw. empirische Forschungsprogramme.<sup>2</sup>

### **3.2.1.2 Ausgestaltung eines prüfungstheoretischen Forschungsprogramms**

#### **3.2.1.2.1 Vorüberlegungen**

##### **3.2.1.2.1.1 Formal-analytische Forschungsprogramme**

Bevorzugte Ansätze zur Ausgestaltung des formal-analytischen Forschungsprogramms stellen der im Wesentlichen auf WYSOCKI zurückzuführende messtheoretische Ansatz<sup>3</sup> sowie neuerdings auch spiel- und agencytheoretische Ansätze<sup>4</sup> dar.

Nach dem *messtheoretischen Ansatz* werden die Ergebnisse der formalen Messtheorie auf die prüferische Urteilsbildung übertragen.<sup>5</sup> „Prüfungen werden als Vergleichsvorgänge zwischen Ist-Gestaltungen und den entsprechenden Soll-Gestaltungen begriffen, die zu Urteilen über die Fehlerhaftigkeit/Fehlerfreiheit des Prüfungsgegenstands führen.“<sup>6</sup> Da jedoch der Prüfungsprozess durch eine hohe Komplexität gekennzeichnet ist, reicht die alleinige Betrachtung von Soll-Ist-Vergleichen nicht aus. Vielmehr werden viele Einzelurteile zu einem Gesamturteil aggregiert. Dabei unterstellt der messtheoreti-

---

<sup>1</sup> Für einen ausführlichen Überblick über Forschungsergebnisse, die eine Handlungsempfehlung hinsichtlich der Normierung oder zumindest eine geeignete Grundlage für die Gestaltung der Prüfungsnormen bieten, sei auf RUHNKE (2000), S. 288-433 verwiesen.

<sup>2</sup> Siehe hierzu ausführlich Kap. 3.1.3.

<sup>3</sup> Vgl. WYSOCKI (2002), Sp. 1886-1899.

<sup>4</sup> Vgl. EWERT (2002), S. 1908-1923.

<sup>5</sup> Vgl. FISCHER-WINKELMANN (1975), S. 127-128.

<sup>6</sup> WYSOCKI (1993), S. 909.

sche Ansatz, dass die Gesamturteile analytisch aus der Summe der jeweiligen Einzelurteile abzuleiten sind.<sup>1</sup> Der messtheoretische Ansatz sah sich in der wissenschaftlichen Diskussion jedoch starker Kritik ausgesetzt. Allem voran wurde bemängelt, dass der Prüfungsprozess ähnlich einem technischen Produktionsprozess als sehr mechanisch aufgefasst und vom individuellen Prüferverhalten sowie vom Gruppenverhalten der Prüfer sowie Geprüften abstrahiert werde.<sup>2</sup> Demnach ist es durchaus gerechtfertigt, Ansätze mit einem höheren Realitätsbezug in die Analyse einzubeziehen.<sup>3</sup>

*Spiel- bzw. agencytheoretische Ansätze*<sup>4, 5</sup> rücken mehrseitige Entscheidungssituationen in den Mittelpunkt der Betrachtung. Im Prüfungskontext geht es zumeist um das Verhalten von Prüfern und Mandanten unter Berücksichtigung von Rechnungslegungs- und Prüfungsnormen.<sup>6</sup> Dabei hängt das Ergebnis sowohl von dem Verhalten eines Spielers (Prüfer) als auch des Gegenspielers (Mandant) ab.<sup>7, 8</sup> Das Hauptanliegen dieser Ansätze besteht darin, das dem Prüfungsprozess innewohnende Konfliktpotenzial zwischen Prüfer und Geprüften explizit zu modellieren.<sup>9</sup> Der Prüfer steht dabei als Garant für die Qualität des Jahresabschlusses, die für eine effiziente Kapitalallokation an den Kapitalmärkten vorausgesetzt wird. Ihm obliegt es insbesondere, die Anteilseigner eines Unternehmens davor zu schützen, dass die Unternehmensleitung ihren Informationsvorteil etwa egoistisch ausnutzt.<sup>10</sup> Neben diesen Erklärungsansätzen zu den Konfliktpotenzialen zwischen den am Prüfungsprozess beteiligten Personen bieten spiel- und agencytheoretische Ansätze auch Begründungen zur Rolle und Bedeutung der Jahresabschlussprüfung im wirtschaftlichen Kontext.<sup>11</sup>

Inwiefern spiel- und agencytheoretische Ansätze für die Entwicklung einer Prüfungstheorie dienlich sein können, ist differenziert zu beurteilen. Einerseits werden sie wegen

---

<sup>1</sup> Vgl. WYSOCKI (1993), S. 909-910.

<sup>2</sup> Vgl. EGNER (1992), S. 1566.

<sup>3</sup> So auch EGNER (1992), Sp. 1567.

<sup>4</sup> Spiel- und agencytheoretische Modelle weisen eine unmittelbare Nähe auf. Agency-Situationen bezeichnet man in der Spieltheorie als ein nicht kooperatives Spiel mit unvollständigen Informationen. Vgl. RIECK (1993), S. 226.

<sup>5</sup> Dem agencytheoretischen Ansatz ist eine höhere Prüfungsspezifität zuzusprechen, da er die vertraglichen Beziehungen zwischen einem Auftraggeber/Geprüften (sog. Prinzipal) und Beauftragten/Prüfer (sog. Agent) zugrunde legt und dabei von einer asymmetrischen Informationsverteilung ausgeht. Vgl. RUHNKE (2000), S. 205.

<sup>6</sup> Vgl. RUHNKE (2000), S. 202 m.w.N. Im Rahmen des Prüfungsprozesses wäre jedoch auch eine mehrseitige Beziehung denkbar, etwa zwischen dem Prüfer, dem Mandanten und den Investoren.

<sup>7</sup> Vgl. RUHNKE (2000), S. 203.

<sup>8</sup> Die Jahresabschlussprüfung als Gegenstand von spieltheoretischen Analysen findet sich bereits bei MERTZ (1963), LOITLSBERGER (1968), und KLAGES (1968).

<sup>9</sup> Vgl. SCHREIBER (2000), S. 15-16.

<sup>10</sup> Vgl. HALLER (1992), S. 414.

<sup>11</sup> Vgl. HALLER (1992), S. 414.



ihrer realitätsfernen Modellierung der Prüfungspraxis stark kritisiert.<sup>1, 2</sup> Andererseits wird argumentiert, dass bereits das Erkennen von Ursache-Wirkungszusammenhängen in abstrakten Modellen oftmals zum Verstehen der Wirklichkeit beiträgt, da „unrealistische“ Modelle weniger zentrale Kausalitäten ausblenden.<sup>3</sup> Zur Veranschaulichung verweist ELSCHEN auf das alltägliche Beispiel einer Straßenkarte. Kein Autofahrer würde von ihr verlangen, dass sie die Realität exakt widerspiegelt. Eine realitätstreue Wiedergabe würde dem Autofahrer das Zurechtfinden vielmehr erschweren als es erleichtern.<sup>4</sup> Aus diesem Blickwinkel blenden Modelle jene Aspekte der Realität aus, die für die aus den Modellen abgeleiteten Aussagen weniger relevant sind.

Von daher scheinen spiel- und agencytheoretische Ansätze durchaus einen Beitrag zur Entwicklung einer Prüfungstheorie leisten zu können. Bei genauerer Betrachtung fällt jedoch auf, dass ihr Nutzen in zweifacher Hinsicht begrenzt ist. Erstens blenden die Modelle fast ausnahmslos die Aufdeckung von unbewussten Fehlern aus.<sup>5</sup> Die Modellprämissen betrachten die nachgebildete Prüfungssituation fast ausschließlich als eine Unterschlagungsprüfung.<sup>6</sup> Unterschlagungsprüfungen stellen jedoch nur einen Teilbereich der Prüfungstätigkeit des Berufsstands dar. Zudem weichen organisatorische und technische Rahmenbedingungen sowie individuelles Verhalten sowohl auf der Prüfer- als auch auf der Geprüfenseite erheblich von „gewöhnlichen“ Jahresabschlussprüfungen ab. Folglich dürfen mittels spiel- und agencytheoretischer Ansätze gewonnene Erkenntnisse nicht per se auf Jahresabschlussprüfungen übertragen werden.

Zweitens werden mit spiel- und agencytheoretischen Ansätzen Prüfungsprozesse aus einem einseitigen strategischen Fokus analysiert. Der Einsatz von Prüfungshandlungen sowie prüferische Urteilsbildungsprozesse über die Normenkonformität von Prüffeldern werden nicht weiter betrachtet, da diese prüfungsbezogenen Problembereiche kaum da-

---

<sup>1</sup> Vgl. etwa GWILLIAM (1987), S. 52.

<sup>2</sup> Beispielhaft lässt sich etwa aufführen, dass (1) die am Prüfungsprozess beteiligten Personen häufig nicht über den angenommenen Informationsstand verfügen, (2) die Abschlussprüfung im Gegensatz zur häufig unterstellten Annahme kein endliches Spiel ist, sondern beliebig häufig wiederholt werden kann und (3) die Modelle sich oftmals auf ein Zwei-Personen-Spiel (Prüfer und Geprüfter) beschränken, die Prüfungspraxis hingegen ein Multi-Personen-Spiel zulässt, was eine Bildung von Interessensgemeinschaften zumindest nicht ausschließt. Auch die Annahme, dass der Prüfer über eine perfekte Prüfungstechnologie verfüge, deren Einsatz es dem Prüfer ermöglicht, sämtliche Fehler beim Mandanten aufzudecken, ist im Hinblick auf die kriminellen Aktivitäten im Rahmen von Bilanzmanipulationen gerade nicht der Fall. Vgl. RUHNKE (2000), S. 215-217 m.w.N.

<sup>3</sup> Vgl. RUHNKE (2000), S. 219 m.w.N.

<sup>4</sup> Vgl. ELSCHEN (1991), S. 1009.

<sup>5</sup> Vgl. RUHNKE (2000), S. 216 m.w.N.

<sup>6</sup> Vgl. etwa TERLINDE (2005), oder FINLEY (1994a), FINLEY (1994b), MATSUMURA/TUCKER (1992) sowie MORTON (1993). Eine umfassenden Übersicht über die spiel- bzw. agencytheoretisch geprägte Forschung im Prüfungswesen gibt RUHNKE (2000), S. 206-215.

zu geeignet sind, als Gegenstand von spiel- und agencytheoretischen Modellbildungen herangezogen zu werden.<sup>1</sup> Es bleibt also festzuhalten, dass diese Ansätze nur partiell einen Nutzen zur Entwicklung einer Prüfungstheorie stiften können.

Zusammenfassend ist zu konstatieren, dass der Rückgriff auf formal-analytische Ansätze zur Theoriebildung im Prüfungswesen in zweifacher Hinsicht beschränkt ist. Die erste Einschränkung ergibt sich aus der unvollständigen Abdeckung der Gesamtheit des realen Phänomens *Prüfung*. Die zweite Einschränkung resultiert aus der mangelnden Abbildung ihrer Komplexität. Letztgenanntes hängt insbesondere mit den stark vom Individualverhalten beeinflussten Urteilsbildungsprozessen des Prüfers zusammen.

Dem ersten Kritikpunkt kann jedoch entschieden entgegengehalten werden, dass der in der wissenschaftlichen Diskussion häufig geforderte Pluralismus der Forschungsprogramme gar nicht vorsieht, dass eine Prüfungstheorie ausschließlich auf der Grundlage eines Ansatzes zu entwickeln ist.<sup>2</sup> Die realitätsfernen Modellannahmen sowie die fehlende Berücksichtigung des prüferischen Urteilsbildungsprozesses bei spiel- und agencytheoretischen Ansätzen einerseits sowie die unrealistische Betrachtung des individuellen Prüferverhaltens im Rahmen des messtheoretischen Ansatzes andererseits können also dadurch kompensiert werden, dass sich die anderen prüfungsbezogenen Bereiche adäquater durch andere Forschungsprogramme erfassen und analysieren lassen. Daher wird im Folgenden die zweite Kategorie von Forschungsprogrammen, namentlich das erfahrungswissenschaftliche bzw. empirische Programm, daraufhin untersucht, inwiefern diese auf dem Weg zur Entwicklung einer Prüfungstheorie eine geeignete Abbildung und Analyse weiterer Bereiche des Prüfungsprozesses gewährleisten kann.

### **3.2.1.2.1.2 Erfahrungswissenschaftliche Forschungsprogramme**

Den größten Anklang zur Ausgestaltung des erfahrungswissenschaftlichen Forschungsprogramms haben in der Prüfungsliteratur der empirisch-kognitive Ansatz<sup>3</sup> sowie der verhaltensorientierte Ansatz<sup>4</sup> gefunden.<sup>5</sup>

---

<sup>1</sup> Ähnlich argumentiert RUHNKE (2000), S. 216.

<sup>2</sup> Vgl. etwa ZÜND (1982), S. 82; WYSOCKI (1993), S. 912; SCHREIBER (2000), S. 16 m.w.N.; RUHNKE (2000), S. 265; RICHTER (1999), S. 269.

<sup>3</sup> Vgl. etwa FISCHER-WINKELMANN (1992), Sp. 1532-1544.

<sup>4</sup> Vgl. EGNER (1992), Sp. 1566-1578 sowie LENZ (2002), Sp. 1924-1938.

<sup>5</sup> Darüber hinaus ließe sich zudem etwa wie bei RUHNKE (2000), S. 229-230 der heuristische Ansatz nach GANS als erfahrungswissenschaftliches Forschungsprogramm aufführen. Da seine wesentlichen

Zielsetzung des *empirisch-kognitiven Ansatzes* ist die Gewinnung von empirisch abgesicherten Aussagen über die in der Realität vorherrschenden Ursache-Wirkungszusammenhänge.<sup>1</sup> Aufgestellte Hypothesen sollen anhand empirisch gewonnener Daten überprüft und gegebenenfalls falsifiziert werden, um so zu einem Bestand an gesicherten allgemeinen Aussagen über die Wirklichkeit zu gelangen. Eine derart ausgerichtete Forschung wird auch als *Prüfstrategie* bezeichnet.<sup>2</sup> Zu Systemen zusammengefasste empirisch gesicherte Aussagen bilden das Kernstück einer empirisch-kognitiv verstandenen Prüfungstheorie. Sie sollen die Verhaltensweisen des Prüfers erklären und für die Zukunft vorhersehbar machen.<sup>3</sup> Demzufolge erscheint es nur konsequent, dass sich der empirisch-kognitive Ansatz um die verstärkte Integration verhaltenswissenschaftlicher Erkenntnisse bemüht.<sup>4</sup>

Charakteristisch für den empirisch-kognitiven Ansatz ist seine strenge Orientierung an den Spielregeln des Kritischen Rationalismus nach POPPER.<sup>5</sup> Zu dessen wesentlichen methodologischen Grundideen gehören zum einen der naturwissenschaftliche Ansatz und zum anderen die Idee eines konsequenten Fallibilismus.<sup>6, 7</sup> Der *konsequente Fallibilismus* erhebt „die These von der prinzipiellen Fehlbarkeit der menschlichen Vernunft“<sup>8</sup>, d.h. theoretische Aussagen sind als vorübergehend anzusehen und revidierbar zu halten. Dem *naturwissenschaftlichen Ansatz* zufolge wird konstatiert, dass alle sozialen und kulturellen Phänomene und somit auch das Prüfungswesen von Gesetzmäßigkeiten beherrscht werden, die in nomologischen Hypothesen (Gesetzesaussagen)<sup>1</sup> erfasst werden können. Nomologische Hypothesen stellen dabei generelle Aussagen dar, welche bestimmte strukturelle Invarianzen beschreiben, die mittels Wenn-Dann- bzw. Je-Desto-

---

Elemente jedoch auf dem Informationsverarbeitungsansatz aus der Kognitionspsychologie basieren, weist dieser Ansatz eine große Ähnlichkeit zum verhaltensorientierten Ansatz auf und wird daher nicht weiter betrachtet. Für eine ausführliche Darstellung des heuristischen Ansatzes soll auf GANS (1986), S. 293-542 verwiesen werden.

<sup>1</sup> Vgl. GROCHLA (1978), S. 79.

<sup>2</sup> Vgl. KUBICEK (1977), S. 6.

<sup>3</sup> Vgl. DREXL (1990), S. 25.

<sup>4</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1539-1540.

<sup>5</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1539-1540; EGNER (1992), Sp. 1566-1567; LENZ (2002), Sp. 1924-1927.

<sup>6</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1532-1535.

<sup>7</sup> Eine weitere Annahme der kritisch-rationalistischen Wissenschaftstheorie ist die Vorstellung einer objektiven, d.h. vom Bewusstsein des empirischen Forschers unabhängigen, jedoch der Erkenntnis zugänglichen Prüfungsrealität. Da eine ernsthaft gemeinte Erörterung dieses Aspekts zwangsläufig auf die Diskussion der den unterschiedlichen wissenschaftstheoretischen Auffassungen zugrunde liegenden philosophischen Grundgedanken führen und weit über den Rahmen der vorliegenden Arbeit hinausgehen würde, wird diese Annahme hier nicht weiter thematisiert. Für eine Darstellung der in diesem Zusammenhang zentralen Begriffe der *Ontologie* sowie *Epistemologie* soll auf SCHÜTTE (1999), S. 211-242 verwiesen werden.

<sup>8</sup> FISCHER-WINKELMANN (1992), Sp. 1533.

Behauptungen versuchen, Phänomene eines bestimmten Objekts zu erfassen und auf diese Weise die Realität zu beschreiben.<sup>2</sup> Von nomologischen Hypothesen wird eine Allgemeinheit verlangt, „die eine essentielle Einschränkung auf ein bestimmtes Raum-Zeit-Gebiet ausschließt“<sup>3</sup> (sog. Postulat der Zeit-Raum-Invarianz).

Nach POPPER gilt eine Hypothese als empirisch bewährt, wenn sie mit ernsthaften und strengen Falsifikationsversuchen konfrontiert und nicht widerlegt wurde und zudem nicht durch wissenschaftlichen Fortschritt überholt wurde.<sup>4</sup> Dieses setzt unabdingbar voraus, dass aufgestellte Hypothesen falsifizierbar sein müssen, d.h. sie müssen mittels empirischer Daten aus der Realität überprüft werden können.<sup>5</sup>

Der empirisch-kognitive Ansatz wird in der prüfungstheoretischen Diskussion jedoch zunehmend kritisiert und als kaum konkretisierbar erachtet. Wesentlicher Kritikpunkt ist dabei der im Zusammenhang mit dem naturwissenschaftlichen Ansatz zu sehende Anspruch auf die Herausbildung nomologischer Hypothesen. Dieses wird mit großer Übereinkunft als unmöglich erachtet.<sup>6</sup> Der Jahresabschlussprüfung ist ein enger Zeit-Raum-Bezug immanent. Es ist nicht von der Hand zu weisen, dass Prüfer unterschiedlicher Länder schon bereits aufgrund der regionalen Spezifität der Rechnungslegungs- und Prüfungsnormen unterschiedliche Vorgehensweisen offenbaren. Hinzu kommt, dass diese Normen einer kontinuierlichen Veränderung unterliegen.<sup>7</sup> Auch die Dynamik im IT-Bereich lässt vermuten, dass der Prüfer seine prüferische Vorgehensweise den Veränderungen im Zeitablauf anpasst. Im Prüfungsbereich muss das für nomologische Hypothesen geforderte Postulat der Zeit-Raum-Invarianz zwangsläufig als gescheitert anerkannt werden.

Sieht man einmal von der Verletzung des Zeit-Raum-Postulats ab,<sup>8</sup> so erweist sich der naturwissenschaftliche Ansatz auch aufgrund der hohen Komplexität des Phänomens Jahresabschlussprüfung als ungeeignet.<sup>9</sup> Während ein Naturwissenschaftler die physika-

---

<sup>1</sup> Siehe auch Kap. 3.1.1.

<sup>2</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1533.

<sup>3</sup> ALBERT (1964), S. 23.

<sup>4</sup> Vgl. LENZ (2002), Sp. 1925.

<sup>5</sup> Vgl. FISCHER-WINKELMANN, Sp. 1534.

<sup>6</sup> Vgl. etwa. RUHNKE (2000), S. 257-258; SCHREIBER (2000), S. 28-29; LENZ (2002), Sp. 1924-1926; DREXL (1990), S. 26; MARTEN/QUICK/RUHNKE (2007), S. 52-53.

<sup>7</sup> Vgl. RUHNKE (2006), S. 651 sowie MARTEN/QUICK/RUHNKE (2007), S. 45-46.

<sup>8</sup> In Anbetracht der harschen Kritik öffnete FISCHER-WINKELMANN (1992), Sp. 1535 das Postulat der Zeit-Raum-Invarianz dahingehend, dass er eine Beschränkung auf einen bestimmten Raum, etwa Deutschland, nicht ausschließt (sog. Quasi-Theorie), sofern ständig deren Nomologisierung vorangetrieben wird.

<sup>9</sup> Vgl. RUHNKE (2006), S. 652 sowie MARTEN/QUICK/RUHNKE (2007), S. 45-46.

lischen Gesetze oder chemische Reaktionen, kurzum die Realität, ohne größere Probleme im Labor nachbilden kann, werden prüfungsbezogene experimentelle Forschungsaktivitäten – die aus Praktikabilitätsgründen zweifelsohne das zentrale Instrument der Prüfungsforschung darstellen, da eine Beobachtung im realen Prüfungsprozess schier unmöglich erscheint – nie die Komplexität einer realen Prüfungssituation im Labor abbilden können.<sup>1</sup> Aus diesen Gründen ist zumindest für den Prüfungsbereich uneingeschränkt anzunehmen, dass Forscher ihre aufgestellten Hypothesen nicht mit der Realität konfrontieren bzw. falsifizieren können und eine streng am Kritischen Rationalismus orientierte Entwicklung einer Prüfungstheorie mithin kaum umsetzbar ist.<sup>2</sup> Dennoch soll der empirisch-kognitive Ansatz nicht gänzlich verworfen werden. Zum einen stellt aus methodologischer Sicht insbesondere die Verwendung einer Prüfstrategie und zum anderen ihr klares Bekenntnis zu einer verhaltenswissenschaftlichen Orientierung einen äußerst fruchtbaren Gedanken für eine realwissenschaftlich geprägte Prüfungstheorie dar.<sup>3</sup> Daher werden im folgenden Abschnitt die Grundlagen der verhaltensorientierten Ansätze skizziert und anschließend ihr Nutzen für eine Realtheorie des Prüfungswesens kritisch überprüft.

Der zentrale Gedanke des *verhaltensorientierten Ansatzes* liegt darin, dass zur Entwicklung einer Prüfungstheorie verstärkt auf das äußere Individual- und/oder Gruppenverhalten abgestellt wird.<sup>4</sup> Er betrachtet das prüferische Verhalten lediglich als einen Spezialfall des menschlichen Verhaltens. Daher werden die in den Verhaltenswissenschaften Psychologie, Sozialpsychologie und Soziologie gesicherten Erkenntnisse über das menschliche Verhalten auch auf den Prüfungsprozess bezogen und untersucht, ob diese bzw. welche von diesen dazu geeignet sind, das individuelle Prüferverhalten in unterschiedlichen Prüfungssituationen zu erklären.<sup>5</sup>

Diese „Konzeptionalisierung des Prüfungswesens vom Individuum her“<sup>6</sup> ist auch deshalb schon gerechtfertigt, weil es sich bei der Jahresabschlussprüfung um ein komplexes und vielschichtiges Phänomen handelt, welches aufgrund fehlender präziser und konkreter Vorschriften über den zu prüfenden Sachverhalt und das prüferische Vorge-

---

<sup>1</sup> Ähnlich argumentiert SCHREIBER (2000), S. 28 -29 m.w.N.

<sup>2</sup> Weitere Kritikpunkte an der rigorosen Umsetzung der Leitideen des Kritischen Rationalismus auf das Prüfungswesen finden sich bei LENZ (2002), Sp. 1924-1926.

<sup>3</sup> Ähnliche Auffassungen finden sich bei RUHNKE (2000), S. 223 sowie SCHREIBER (2000), S. 29.

<sup>4</sup> Vgl. RUHNKE (2000), S. 224.

<sup>5</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1539-1540; LENZ (2002), Sp. 1926.

<sup>6</sup> RUHNKE (2000), S. 224.

hen nur schwierig standardisiert werden kann.<sup>1</sup> Die Jahresabschlussprüfung als Phänomen „an sich“ eignet sich folglich nur wenig, um als Gegenstand der Theoriebildung herangezogen zu werden. Vielmehr sollen Prüfungsprozesse als soziales Geschehen betrachtet werden, an denen unterschiedliche Individuen beteiligt sind (Prüfungsassistenten, Prüfungsleiter, Manager, Geprüfte usw.). Soziales Geschehen wird dabei als komplexes Ergebnis aus dem Zusammenspiel individueller Handlungen der beteiligten Personen betrachtet.<sup>2</sup> Folglich rücken Aussagen über emotionale, kognitive und motivationale Prozesse in den Mittelpunkt der Theoriebildung.<sup>3</sup> Eine als Sozialwissenschaft verstandene Prüfungslehre hat es dabei zur Aufgabe, Erklärungen für Phänomene zu finden, die innerhalb eines bestimmten situativen Kontexts anzutreffen sind.<sup>4</sup>

Der verhaltensorientierte Ansatz sieht sich in der Prüfungsliteratur jedoch auch einiger Kritik ausgesetzt.<sup>5</sup> Hauptgegenstand der kontroversen Diskussion ist dabei die interdisziplinäre Ausrichtung des verhaltensorientierten Ansatzes.<sup>6</sup> Problematisch sei dabei insbesondere, welche verhaltenswissenschaftlichen Erkenntnisse zur Erklärung menschlichen Verhaltens auf welche prüfungsspezifischen Sachverhalte übertragen werden sollen, zumal selbst in den Verhaltenswissenschaften nur Teiltheorien allgemeinemenschlichen Verhaltens existieren (sog. Auswahlproblem).<sup>7</sup> Als Höhepunkt dieser Kontroverse kann der „Dilettantismusvorwurf“ an eine grenzüberschreitende Arbeit zwischen der Betriebswirtschafts- bzw. Prüfungslehre und den Verhaltenswissenschaften verzeichnet werden.

Diese Kritik wurde in der Prüfungsliteratur jedoch vehement zurückgewiesen.<sup>8</sup> EGNER entgegnet zu Recht „warum denn die vielen grenzüberschreitenden Arbeiten zwischen Prüfungslehre und Recht und Mathematik und Statistik nicht zu Dilettantismus verführen, wohl aber grenzüberschreitende Arbeiten zwischen Betriebswirtschaftslehre und Psychologie und/oder Soziologie?“.<sup>9</sup> Ferner wird entgegengehalten, dass eine strikte Trennung von Disziplinen vielmehr zu einer Zerschneidung von Kommunikationsbe-

---

<sup>1</sup> Vgl. EGNER (1992), Sp. 1567.

<sup>2</sup> Vgl. SCHANZ (1993), Sp. 4523. Ähnlich ALBERT (1977), S. 183.

<sup>3</sup> Vgl. LENZ (2002), S. 1927.

<sup>4</sup> Vgl. SCHANZ (1979), S. 91.

<sup>5</sup> Vgl. beispielhaft FISCHER-WINKELMANN (1992), Sp. 1540.

<sup>6</sup> Vgl. hierzu die Diskussion zwischen SCHNEIDER (1981) bzw. SCHNEIDER (1984) und EGNER (1984).

<sup>7</sup> Vgl. FISCHER-WINKELMANN (1992), Sp. 1540.

<sup>8</sup> Vgl. stellvertretend SCHREIBER (2000), S. 16, 52-56 sowie EGNER (1992), Sp. 1575-1576. Auch LÜCK (1991), S. 289, GANS (1986), S. 50, RUHNKE (2000), S. 259, 265 und RICHTER (1999), S. 269 halten weiterhin an der Interdisziplinarität fest.

<sup>9</sup> EGNER (1992), Sp. 1575-1576.

ziehungen, einer einseitigen Sichtweise und Verödung der Grenzbereiche zwischen den unterschiedlichen Disziplinen führe.<sup>1,2</sup>

Weniger Vorbehalte gegen eine interdisziplinäre Ausrichtung herrschen hingegen in der angloamerikanischen Prüfungsforschung.<sup>3</sup> Hier wurde seit Mitte der siebziger Jahre eine kaum noch überschaubare Vielzahl von interdisziplinär ausgerichteten Forschungsergebnissen veröffentlicht.<sup>4</sup> Bei einer nicht unwesentlichen Anzahl von Publikationen wirkten sogar namhafte Psychologieforscher mit,<sup>5</sup> was dem o.a. Auswahlproblem eines verhaltensorientierten Ansatzes sicherlich entgegenwirken kann.

Auch der Autor spricht sich eindeutig für eine interdisziplinär ausgerichtete Prüfungsforschung aus. Schließlich hat eine als angewandte Wissenschaft verstandene Prüfungslehre ihre zentrale Aufgabe darin zu sehen, die Bedürfnisse und Probleme der Prüfungspraxis zu berücksichtigen. CUSHING bemerkt diesbezüglich zutreffend: „rather than choose our methodology first and our problem second, we must choose the problem first and then adopt the research methods most appropriate for the problem“.<sup>6</sup> Da Prüfungsprozesse nun einmal zu einem nicht unwesentlichen Teil aus Handlungen und Interaktionen zwischen Prüfern und Geprüften bestehen, könnte eine Nicht-Berücksichtigung verhaltenswissenschaftlicher Aspekte zu einem realitätsfernen Unterfangen führen. Jeder Versuch, die Realität zu erfassen bzw. zu gestalten, ohne dabei die in der Realität vorherrschenden charakteristischen Gegebenheiten in die Forschungsbemühungen weitestgehend zu integrieren, nährt nach Ansicht des Autors das häufig anzutreffende Vorurteil, dass Theorie doch nur herzlich wenig mit der Praxis zu tun habe.

### **3.2.1.2.1.3 Würdigung der bisherigen Überlegungen**

Resümierend bleibt demnach festzuhalten, dass vieles dafür spricht, eine Prüfungstheorie maßgeblich nach den Inhalten und Methoden eines erfahrungswissenschaftlichen

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 52.

<sup>2</sup> Beispiele für eine verhaltensorientierte und somit interdisziplinäre Ausrichtung in der betriebswirtschaftlichen Forschung finden sich etwa in den Bereichen Controlling und Marketing. Für Erstgenanntes siehe etwa SCHORB (1994), für Letzteres bspw. KROEBER-RIEL/WEINBERG (2003) sowie BACKHAUS (2003).

<sup>3</sup> Vgl. WEIBLER (1996), S. 658.

<sup>4</sup> Vgl. RICHTER (1997a), S. 20. Eine umfassende Darstellung der Ergebnisse empirischer Prüfungsforschung im angloamerikanischen Raum gibt RUHNKE (2000), S. 288-429. Eine rein auf das Informationsverhalten im Rahmen der Prüfungsdurchführung gerichtete Analyse findet sich hingegen bei SCHREIBER (2000), S. 112-177.

<sup>5</sup> Vgl. etwa EINHORN (1976); SHANTEAU (1989); HOGARTH (1991).

<sup>6</sup> CUSHING (1977), S. 7.

bzw. empirischen Forschungsprogramms auszurichten. Dieses soll nicht bedeuten, dass formal-analytische bzw. entscheidungslogische Forschung völlig nutzlos sei. Analytisch-theoretische Arbeit hat auch in einer als empirische oder Erfahrungswissenschaft aufgefassten Prüfungslehre ihren Platz.<sup>1</sup> Der Wert formaler Modelle könnte vor allem darin gesehen werden, grundlegende Erkenntnisse über Ergebnis- und Lösungsstrukturen zu generieren.<sup>2</sup> Dabei könnten Sie als Orientierungshilfe dienen und eine allgemeine Grundlage für die Diskussionen über Ursache-Wirkungszusammenhänge darstellen.<sup>3</sup> *Ruhnke* attestiert formalen Modellen ferner eine *empirische Komponente*, sofern sie für empirische Forschungsaktivitäten anschlussfähig sind, d.h. eine geeignete Basis für empirisch zu überprüfende Hypothesen bilden, und/oder sie als Analyseinstrument für bereits existierendes empirisches Wissen herangezogen werden, d.h. vorhandene empirische Befunde über Ursache-Wirkungszusammenhänge modelliert und analysiert werden.<sup>1</sup>

Wie nun aber soll der Prüfungsforscher konkret vorgehen, um empirische Erkenntnisse über das Prüfungswesen zu erlangen? Es stellt sich also im Zuge eines erfahrungswissenschaftlich geprägten Wissenschaftsverständnisses die Frage, wie ein darauf gerichtetes Forschungsprogramm inhaltlich und methodisch ausgestaltet werden soll. Zur Beantwortung dieser Frage wird im Folgenden ausschließlich auf die beiden bereits vorgestellten Ansätze (empirisch-kognitiver und verhaltensorientierter Ansatz) zurückgegriffen, die im Übrigen aufgrund ihres Bekenntnisses zur Verhaltenswissenschaft einen engen Bezug zueinander aufweisen.

Sieht man einmal von der Unmöglichkeit nomologischer Hypothesen im Prüfungsbereich zum einen und ihrer mangelnden Falsifizierbarkeit in der Prüfungsrealität zum anderen ab, sind dem empirisch-kognitiven Ansatz aufgrund der rigorosen Anwendung des Kritischen Rationalismus erhebliche methodologische Vorzüge zu bescheinigen. Die Aufstellung von Hypothesen und deren Falsifikation bzw. Verifikation (sog. Prüfstrategie) soll sich daher – wenn auch nicht in einer dem Kritischen Rationalismus entsprechenden Strenge – in dem angestrebten empirischen Forschungsprogramm wieder finden. Allerdings sind pragmatische Ersatzkriterien zur Aufstellung und Verwerfung von Hypothesen dahingehend zu akzeptieren, dass die „Illusion“ einer Prüfungstheorie im Sinne eines Systems aufeinander bezogener nomologischer Hypothesen zugunsten

---

<sup>1</sup> Vgl. EGNER (1992), Sp. 1567.

<sup>2</sup> Vgl. EWERT (1993), S. 722.

<sup>3</sup> Vgl. LAUX (1990), S. 25-26.



von isolierten und miniaturisierten Theorieteilen ohne hohe Allgemeingültigkeit und gegenseitiger Konsistenz aufzugeben ist.<sup>2</sup>

Methodologisch weniger elaboriert erweist sich hingegen der verhaltenswissenschaftliche Ansatz. Der erhebliche Nutzen dieses eher als pragmatisch zu bezeichnenden Ansatzes liegt in erster Linie darin, dass er als wertvoller „Ideengeber“ für die empirischen Forschungsaktivitäten im Prüfungswesen fungiert, indem er die reichhaltigen Erkenntnisse der Psychologie und Soziologie für den Prüfungsbereich zugänglich macht. Die empirische Prüfungsforschung kann hierdurch in zweifacher Hinsicht profitieren. Einerseits können hieraus entnommene Modelle und Theorien menschlichen Verhaltens zur Erfassung und Beschreibung der prüferischen Vorgehensweise herangezogen werden und andererseits können empirisch gesicherte Verhaltensmuster als Hypothesen übernommen werden, die es im Prüfungskontext zu bestätigen bzw. zu widerlegen gilt.<sup>3</sup>

#### **3.2.1.2.2 Skizzierung eines erfahrungswissenschaftlichen Forschungsprogramms für das Prüfungswesen**

Ein prüfungstheoretisches Forschungsprogramm, welches sich zum einen – soweit möglich – an den Spielregeln des Kritischen Rationalismus orientiert und sich nicht den umfassenden Erkenntnissen der Verhaltenswissenschaften verschließt, und zum anderen versucht, formal-analytisch hergeleitete Ideen zu integrieren, könnte durch folgenden mehrstufigen Prozess wiedergegeben werden.

Im *ersten Schritt* sollte der Prüfer zunächst das zu betrachtende Prüfungsproblem ausführlich darstellen. Im *zweiten Schritt* obliegt es dem Forscher, dieses Problem durch ein geeignetes Modell oder eine Theorie abzubilden. Die Abbildung kann dabei auf formal-analytischen oder verhaltenswissenschaftlichen Elementen basieren. Im *dritten Schritt* ist eine Hypothese zu entwickeln, die eine Vermutung über das noch zu beo-

---

<sup>1</sup> Vgl. RUHNKE (2000), S. 218-219.

<sup>2</sup> Vgl. LENZ (2002), Sp. 1926.

<sup>3</sup> Allerdings ist hierbei auch zu beachten, dass es nicht auf eine unreflektierte und unangepasste Übertragung von Hypothesen, Modellen und Theorien der Verhaltenswissenschaften auf den Prüfungsbereich hinauslaufen soll. Diese sog. „borrow-and-transfer“-Orientierung prägte insbesondere die Anfänge der angloamerikanischen Prüfungsforschung (einen Überblick gibt etwa SOLOMON/SHIELDS (1995), S. 142). Zweckmäßiger erscheint hingegen die sog. „contrast-before-transfer“-Orientierung, bei der Unterschiede zwischen den in der psychologischen Forschung überwiegend anzutreffenden einfachen Aufgabenstellungen und den prüfungsspezifischen Problemstellungen explizit berücksichtigt werden. Diese Vorgehensweise gewann zu Beginn der neunziger Jahre zunehmende Bedeutung in der Prüfungsforschung (Vgl. etwa ASHTON/ASHTON (1995), S. 19-22; SOLOMON/SHIELDS (1995), S. 141; HAYNES (1993), S. 30-37).

bachtende prüferische Verhalten ausspricht. Diese Hypothese kann entweder logisch-deduktiv abgeleitet werden oder eine empirisch abgesicherte Erkenntnis aus den Verhaltenswissenschaften sein (sog. borrow-and-transfer). Anstelle einer Hypothesenentwicklung können unter Umständen auch Forschungsfragen formuliert werden. Dieses geschieht regelmäßig, falls Hypothesen nicht logisch-deduktiv abgeleitet werden können oder keine gesicherten empirischen Erkenntnisse über den zu beurteilenden Sachverhalt vorliegen. Die Formulierung von Forschungsfragen ist häufig im Rahmen von explorativ ausgerichteten Untersuchungen zu finden.<sup>1</sup> Im anschließenden *vierten Schritt* wird das prüferische Verhalten erfasst. Mit diesem Schritt geht regelmäßig die Gewinnung von Daten einher. Im abschließend *fünften Schritt* werden die erlangten Daten ausgewertet und die anfänglich aufgestellten Hypothesen entweder bestätigt oder verworfen, wobei kein Anspruch auf Allgemeingültigkeit oder Konsistenz zu bisherigen Erkenntnissen erhoben werden kann. Wurden anstelle von Hypothesen Forschungsfragen formuliert, so können diese durch die Datenauswertung möglicherweise beantwortet werden. In der folgenden Graphik wird das dargestellte Forschungsprogramm illustriert.

---

<sup>1</sup> Beispiele für ein derartiges Vorgehen finden sich etwa bei BEDARD/BIGGS/MARONEY (1998), ASARE/TROMPETER/WRIGHT (2000), GREEN (2004a), ASARE/WRIGHT (1997), ANDERSON/KAPLAN/RECKERS (1992), BIERSTAKER/BEDARD/BIGGS (1999), GREEN/TROTMAN (2003).

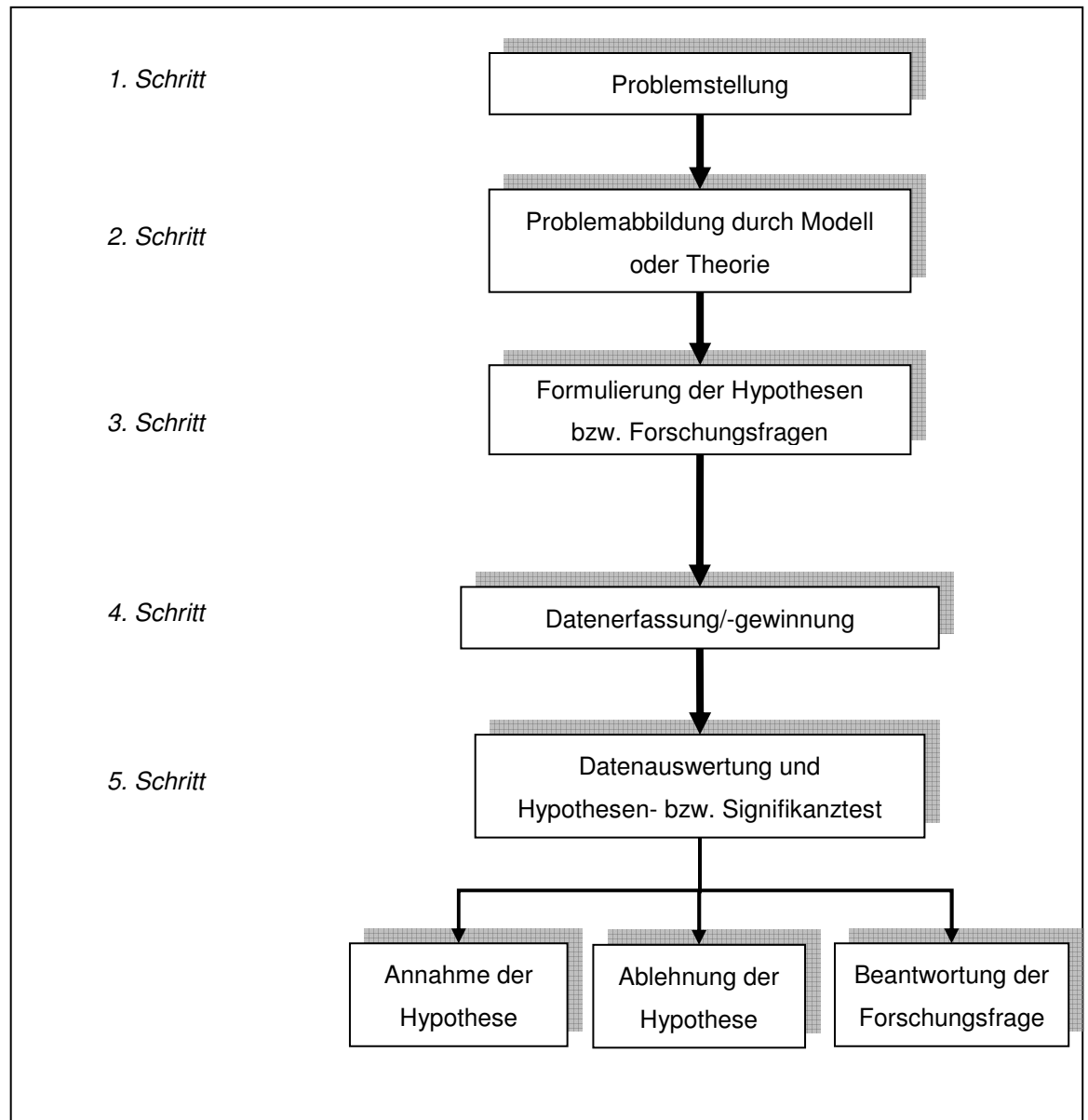


Abbildung 4: Erfahrungswissenschaftliches Forschungsprogramm für das Prüfungswesen

Um das skizzierte empirische Forschungsprogramm vollständig zu validieren, gilt es im Folgenden zwei weitere Fragen zu beantworten. Erstens, welche verhaltensorientierten Modelle kommen zur Abbildung eines prüfungsspezifischen Sachverhalts (Schritt 2) in Betracht? Hier stehen insbesondere Regressions- und Attributionsmodelle, Modelle probabilistischer Wahrscheinlichkeitsschlüsse sowie Problemlösungsmodelle im Mittelpunkt der Betrachtung.<sup>1</sup> Zweitens, welche Methoden zur Datengewinnung (Schritt 4) und Datenauswertung (Schritt 5) stehen dem empirischen Forscher zur Verfügung? Während die Beantwortung der ersten Frage den Gegenstand des folgenden Abschnitts bildet, erfolgt die Beantwortung der zweiten Frage unter Kap. 3.4.

### 3.2.2 Verhaltensorientierte Modelle zur Beschreibung des prüferischen Vorgehens

#### 3.2.2.1 Regressions- und Attributionsmodelle

Die insbesondere zu Beginn der empirischen Prüfungsforschung häufig eingesetzten *Regressions- und Attributionsmodelle*, deren bekanntester Vertreter das *Brunswick'sche Linsenmodell*<sup>2</sup> ist, betrachten den Prüfer als eine „black box“ und versuchen zu erklären, inwieweit bestimmte Informationen (sog. Prädikatoren oder „cues“) für die prüferische Urteilsbildung verantwortlich sind.<sup>3</sup> Es handelt sich dabei um eine reine Input-Output-Betrachtung, bei der versucht wird, die Zusammenhänge zwischen den einzelnen Informationen und dem Prüfungsurteil mathematisch zu modellieren.<sup>4</sup> Charakteristisch für diese Modelle ist, dass sie eine Ermittlung der relativen Wichtigkeit einzelner Informationen für den Urteilsbildungsprozess anstreben.<sup>5</sup> Dies kann etwa anhand von Regressionskoeffizienten geschehen.<sup>6</sup>

Kritisiert wird allerdings, dass *Regressions- und Attributionsmodelle* von einer bloßen Input-Output-Beziehung ausgehen und den Urteilsbildungsprozess an sich vollends ausklammern.<sup>7</sup> Von den individuellen Verhaltensweisen eines Prüfers, etwa in Bezug auf seine Informationsbeschaffungsaktivitäten, wird vollends abstrahiert.<sup>8</sup> Die Modelle umfassen eine als vollständig erachtete Informationsmenge, so dass ein wie in der Prüfungsrealität zu vermutendes „Nachfassen“ des Prüfers, um weitere Informationen zu erhalten, vollends herausdefiniert wird.<sup>9</sup> Eine völlige Instrumentalisierung der prüferischen Urteilsbildung greift allerdings zu kurz. Vielmehr sind weitere wichtige Einflussfaktoren zu berücksichtigen, wie etwa äußere Rahmenbedingungen sowie das Wissen und die kognitiven Fähigkeiten des Prüfers.<sup>10</sup> Mittels Regressions- und Attributionsmodellen lassen sich allenfalls gut strukturierte und weniger komplexe Probleme modellie-

---

<sup>1</sup> Vgl. auch SCHREIBER (2000), S. 12-15; RUHNKE (2000), 235-240; SOLOMON/SHIELDS (1995), S. 152-162.

<sup>2</sup> Eine umfassende Darstellung gibt GANS (1986), S. 66-150.

<sup>3</sup> Vgl. HALLER (1992), S. 417.

<sup>4</sup> Vgl. RUHNKE (2000), S. 235-236 m.w.N. sowie GANS (1986), S. 67.

<sup>5</sup> Vgl. RUHNKE (2000), S. 236 m.w.N. sowie HALLER (1992), S. 417.

<sup>6</sup> Vgl. LENZ (2002), Sp. 1930.

<sup>7</sup> Vgl. ASHTON, R. (1982), S. 13 sowie LIBBY (1981), S. 40.

<sup>8</sup> Vgl. SCHREIBER (2000), S. 12 sowie RUHNKE (2000), S. 237.

<sup>9</sup> Vgl. FIEDLER (1980), S. 50; BIGGS/MOCK (1983), S. 241.

<sup>10</sup> Vgl. SCHREIBER (2000), S. 12. Eine umfassende Übersicht über die personenabhängigen und prüfungsumfeldbedingten Faktoren, denen bei der Analyse der prüferischen Urteilsbildung eine erhöhte Aufmerksamkeit zu widmen ist, findet sich bei SCHREIBER (2000), S. 83-110.

ren, die im Prüfungswesen zumeist nur selten anzutreffen sind.<sup>1</sup> Daher werden diese Modelle im Folgenden nicht weiter betrachtet.

### **3.2.2.2 Modelle probabilistischer Wahrscheinlichkeitsschlüsse**

#### **3.2.2.2.1 Wesen probabilistischer Modelle**

Probabilistische Modelle lösen das Problem der „black box“-Betrachtung, indem sie nicht das Ergebnis des gesamten Bearbeitungsprozesses betrachten, sondern das Ergebnis einzelner Prozessphasen. Darüber hinaus findet in diesem Modell auch das individuelle Vorwissen Berücksichtigung, was ihre Einsetzbarkeit zur realitätsnahen Abbildung von Prüfungsprozessen zudem erhöht.<sup>2</sup>

Modelle probabilistischer Wahrscheinlichkeitsschlüsse beschäftigen sich mit der Schätzung und Anpassung subjektiver Wahrscheinlichkeiten sowie der Erforschung systematischer Verzerrungen im menschlichen Wahrscheinlichkeitsdenken (sog. biases).<sup>3</sup> Dabei geht das Modell davon aus, dass der Prüfer eine seinen prüfungsspezifischen Vorkenntnissen entsprechende subjektive a priori-Fehlerwahrscheinlichkeit abgibt und diese Wahrscheinlichkeitseinschätzung aufgrund der Erlangung weiterer Informationen kontinuierlich modifiziert.<sup>4</sup> Als normative Grundlage für subjektive Wahrscheinlichkeitsschlüsse wird das Bayes-Theorem herangezogen.<sup>5</sup> Allerdings unterliegen Individuen bei ihren Wahrscheinlichkeitsschlüssen gewissen Fehlertendenzen, die sich darin äußern, dass sich ihre Wahrscheinlichkeitseinschätzungen von denen unterscheiden, die sich bei der Verwendung des Bayes-Theorems ergeben.<sup>6</sup>

Zur Ergründung dieser Verhaltensanomalien hat sich eine eigenständige „heuristics-and-biases“-Forschung entwickelt, die sich ausschließlich mit der Erforschung von kognitiven Verzerrungen bzw. systematischen Fehlertendenzen bei Wahrscheinlichkeitsurteilen beschäftigt.<sup>1</sup> Von wesentlicher Bedeutung ist hierbei die Frage, inwiefern die prüferische Urteilsbildung durch die Verwendung von Heuristiken (heuristics) gekenn-

---

<sup>1</sup> Vgl. WRIGHT, W. (1982), S. 71; LARCKER/LESSIG (1983), S. 59; MARTEN/QUICK/RUHNKE (2006), S. 393; SCHREIBER (2000), S. 82; MARTEN/QUICK/RUHNKE (2007), S. 50-51.

<sup>2</sup> Vgl. RUHNKE (2000), S. 239.

<sup>3</sup> Eine detaillierte Darstellung findet sich bei GANS (1986), S. 151-193.

<sup>4</sup> Vgl. RUHNKE (2000), S. 239.

<sup>5</sup> Vgl. stellvertretend GANS (1986), S. 163-164.

<sup>6</sup> Vgl. LENZ (2002), Sp. 1930.

zeichnet ist.<sup>2</sup> Heuristiken sind dabei als vereinfachte Vorgehensweisen (sog. Faustregeln) zur Problembearbeitung zu verstehen, die den mentalen bzw. kognitiven Aufwand reduzieren.<sup>3</sup> Ihr Einsatz ist oft effizient, kann aber auch zu kognitiven Verzerrungen (biases) führen.<sup>4</sup>

### **3.2.2.2.2 Heuristiken bei Wahrscheinlichkeitsurteilen**

#### **3.2.2.2.2.1 Begrenzte Rationalität und Heuristiken**

Modelle *begrenzter Rationalität* beziehen sich auf Beschränkungen der Umwelt, wie etwa Kosten für Informationen, oder auf mentale Beschränkungen, wie etwa ein begrenztes Erinnerungs- bzw. Informationsverarbeitungsvermögen. In Bereichen, die durch eine hohe Komplexität gekennzeichnet sind, hierzu gehört auch das Gebiet der Jahresabschlussprüfung,<sup>5</sup> ist insbesondere letztgenannte Beschränkung zu berücksichtigen, also die begrenzte Kapazität zur mentalen Erfassung von für die Urteilsbildung relevanten Informationen.<sup>6</sup>

In Bereichen, in denen Urteile und Entscheidungen aufgrund ihrer hohen Komplexität nur unter Unsicherheit getroffen werden können, verwenden Individuen u.a. „kognitive Daumenregeln“, um die ihnen gegenüberstehende Komplexität reduzieren und somit handhaben zu können.<sup>7</sup> Solche Vereinfachungsverfahren werden in der Literatur üblicherweise als *Heuristiken* bezeichnet.<sup>8</sup> Heuristiken sind mit einem geringen kognitiven Aufwand verbunden, gehen eher unsystematisch vor und sind von den Erfahrungen des Individuums abhängig. Nach der sog. *Dual-Process-Theory* wird davon ausgegangen, dass Individuen bei bestimmten Sachverhalten häufig auf Heuristiken zurückgreifen und

---

<sup>1</sup> Diese Forschungsrichtung wurde in den siebziger Jahren insbesondere durch die Arbeiten von AMOS TVERSKY und DANIEL KAHNEMAN geprägt. Vgl. etwa KAHNEMAN/TVERSKY (1972), S. 430-454 sowie TVERSKY/KAHNEMAN (1974), S. 1124-1131.

<sup>2</sup> Vgl. LENZ (2002), Sp. 1930.

<sup>3</sup> Vgl. LIBBY (1981), S. 52.

<sup>4</sup> Vgl. LENZ (2002), Sp. 1930.

<sup>5</sup> Vgl. etwa SCHREIBER (2000), S. 161 oder ASHTON/ASHTON (1995), S. 6.

<sup>6</sup> Im Bereich der Jahresabschlussprüfung sind grundsätzlich auch Fragen hinsichtlich der Transaktionskosten von Informationen relevant. Dies wird insbesondere deutlich, wenn man nach RICHTER (1999), S. 280-291 die Jahresabschlussprüfung auch als Konflikthandhabungsprozess charakterisiert, wobei sich insbesondere das mit der Wahl einer Informationsquelle verbundene Konfliktpotenzial auf das Verhalten des Prüfers auswirken kann. Siehe hierzu ausführlich SCHREIBER (2000), S. 107-110. Da hier allerdings verstärkt sozio-emotionale und weniger kognitive Aspekte im Vordergrund stehen, wird hier nicht weiter darauf eingegangen.

<sup>7</sup> Vgl. GANS (1986), S. 166. RICHTER (1999), S. 270-274 bezeichnet den Prüfungsprozess daher u.a. auch als Komplexitätshandhabungsprozess.

<sup>8</sup> Vgl. etwa BIDDLE/JOYCE (1982), S. 165.

bei anderen Sachverhalten systematische Problemlösungsstrategien verwenden.<sup>1</sup> Im Gegensatz zu *Algorithmen*, die entweder zu einer optimalen oder zu keiner Lösung eines Problems führen, kommen Heuristiken regelmäßig zu einer Lösung, können jedoch keine optimale Lösung garantieren.<sup>2</sup> Somit können heuristische Lösungen auch als *Quasi-Lösungen* bezeichnet werden.

Obwohl Heuristiken häufig zu nützlichen Urteilen führen, sind mit ihnen nicht selten systematische Fehleinschätzungen verbunden.<sup>3</sup> Systematische Fehleinschätzungen liegen vor, wenn für die Urteilsbildung relevante Faktoren nicht nur zufällig, sondern konsistent zu stark oder zu schwach berücksichtigt werden.<sup>4</sup> Solche systematischen Abweichungen sollen im Folgenden als *biases* bezeichnet werden.

### 3.2.2.2.2 Arten von Urteilsbildungsheuristiken

In der empirischen Prüfungsforschung werden im Wesentlichen die auf KAHNEMAN/TVERSKY<sup>5</sup> zurückzuführenden Heuristiken diskutiert. Hierbei handelt es sich um die

- Verfügbarkeits-Heuristik,
- Repräsentativitäts-Heuristik und
- Anker-Heuristik.

Die *Verfügbarkeits-Heuristik* wurde von TVERSKY/KAHNEMAN als eine „heuristic for judging frequencies and probabilities“ bezeichnet.<sup>6</sup> Sie argumentieren, dass die Verfügbarkeit einer Information ein wichtiger Hinweis ist, den Individuen benötigen, um Urteile fällen zu können. Ein Ereignis wird umso wahrscheinlicher eingestuft, je leichter und schneller ein Individuum in der Lage ist, sich Beispiele für das Ereignis vorzustellen oder in Erinnerung zurufen.<sup>7</sup> Als Konsequenz dieser Heuristik wird leicht verfügba-

---

<sup>1</sup> Vgl. EPSTEIN/PACINI (1999), S. 462-483.

<sup>2</sup> Ähnlich argumentieren BRANDER/KOMPA/PELTZER (1989), S. 137.

<sup>3</sup> Vgl. TVERSKY/KAHNEMAN (1982), S. 3.

<sup>4</sup> Vgl. KOROBIKIN/ULEN (2000), S. 1085.

<sup>5</sup> Vgl. hierzu grundlegend TVERSKY/KAHNEMAN (1974), S. 1124-1130 und TVERSKY/KAHNEMAN (1979), S. 263-291.

<sup>6</sup> TVERSKY/KAHNEMAN (1973), S. 207.

<sup>7</sup> Vgl. TVERSKY/KAHNEMAN (1982), S. 163-164.

ren Ereignissen eine höhere Wahrscheinlichkeit zugeordnet, d.h. die Einschätzungen orientieren sich nach Extremen statt nach Durchschnittswerten.<sup>1</sup>

Die Verfügbarkeits-Heuristik kann in zahlreichen Situationen eine sinnvolle Regel für die Einschätzung einer Wahrscheinlichkeit darstellen. D.h. Ereignisse, die in der Vergangenheit häufig aufgetreten sind, werden auch weiterhin häufig auftreten. Dennoch gibt es auch andere Faktoren, die sich auf die Verfügbarkeit im Gedächtnis auswirken. So können etwa häufige und einprägsame Berichterstattungen über seltene Ereignisse, wie etwa Flugzeugabstürze, Terroranschläge oder Bilanzskandale, systematische Fehleinschätzungen (*biases*) herbeiführen.<sup>2</sup> Auch die eigenen Erfahrungen, etwa aus vergangenen Prüfungsaufträgen, können zu *biases* führen, da sie nicht repräsentativ für die Grundgesamtheit sein müssen. In der empirischen Prüfungsforschung wurde die Verfügbarkeits-Heuristik häufig im Zusammenhang mit der Generierung von potenziellen Abweichungsursachen untersucht.<sup>3</sup>

Unter einer *Repräsentativitäts-Heuristik* verstehen KAHNEMAN/TVERSKY, dass die subjektive Wahrscheinlichkeit für ein Ereignis umso größer eingeschätzt wird, je hervorstechender das Ereignis für die Population ist, aus der sie stammt.<sup>4</sup> Individuen schätzen folglich typische Ereignisse als wahrscheinlicher ein als untypische Ereignisse.

Diese Repräsentativitäts-Heuristik kann zwar in vielen Situationen zu nützlichen Ergebnissen führen. Da jedoch die Wahrscheinlichkeitsurteile über Ereignisse nicht auf tatsächliche Häufigkeiten, sondern auf deren Repräsentativität basieren, besteht die Gefahr, dass weitere für die Beurteilung von Wahrscheinlichkeiten wichtige Kriterien vernachlässigt werden und mit systematischen Fehleinschätzungen (*biases*) zu rechnen ist.<sup>5</sup> Im Prüfungsbereich konzentriert man sich im Wesentlichen auf zwei mit der Anwendung der Repräsentativitäts-Heuristik verbundene *biases*. Dabei handelt es sich einerseits um den *Konjunktionsfehler* und andererseits um den *Basisratenfehler*.<sup>6</sup>

---

<sup>1</sup> Vgl. BIDDLE/JOYCE (1982), S. 183; LIBBY (1981), S. 65.

<sup>2</sup> Beispielsweise schätzen Probanden die relativen Häufigkeiten von Krankheiten oder Ereignissen, über die viel publiziert wird, wie etwa Krebs oder Tornados, höher ein als die relativen Häufigkeiten von Krankheiten, die weniger diskutiert werden, wie etwa Asthma oder Diabetes. Vgl. SLOVIC/KUNREUTHER/WHITE (1974), S.187-205.

<sup>3</sup> Siehe hierzu insbesondere die Ausführungen zu Kap. 3.5.3.3.1.1.

<sup>4</sup> Vgl. KAHNEMAN/TVERSKY (1972), S. 431.

<sup>5</sup> Vgl. BIDDLE/JOYCE (1982), S. 168.

<sup>6</sup> Vgl. hierzu etwa Kap. 3.5.5.4.3. Weitere in Verbindung mit der Repräsentativitäts-Heuristik stehende *biases* sind die ‚Vernachlässigung des Stichprobenumfangs‘, die ‚falsche Vorstellung von Zufallsmerkmalen‘ und die ‚Außerachtlassung der Regression zur Mitte‘. Vgl. hierzu ausführlich TVERSKY/KAHNEMAN (1974), S. 1124-1127. Da diese im Prüfungsbereich jedoch kaum diskutiert wurden, sollen sie an dieser Stelle nicht weiter berücksichtigt werden.



Müssen Individuen mehrere potenzielle Ereignisse für ein Wahrscheinlichkeitsurteil kombinieren, so kommt es häufig zu einer Überschätzung der Wahrscheinlichkeit für eine Konjunktion.<sup>1</sup> Aus statistischer Sicht ist dies ein Fehler (sog. *Konjunktionsfehler*), da die Wahrscheinlichkeit einer konjunktiven Verknüpfung von Ereignissen nicht größer sein kann als die Wahrscheinlichkeit eines Einzelereignisses, d.h.  $P(A \cap B)$  kann nicht wahrscheinlicher sein als  $P(A)$  oder  $P(B)$ . TVERSKY/KAHNEMAN erklären diese subjektiven Fehleinschätzungen (*biases*) damit, dass die Repräsentativität von multiplen Gründen als höher wahrgenommen wird als die Repräsentativität eines Einzelgrunds.<sup>2</sup> In zahlreichen psychologischen Studien konnte der Konjunktionsfehler unabhängig von den statistischen Kenntnissen der Probanden nachgewiesen werden.<sup>3</sup>

Ignorieren Individuen bei der Beurteilung von Wahrscheinlichkeiten die Grundgesamtheit des zugrunde liegenden Ereignisses, so wird diese systematische Fehleinschätzung (*bias*) als *Basisratenfehler* bezeichnet.<sup>4</sup> Der Fehler der beurteilenden Person ist darin zu sehen, dass nicht berücksichtigt wird, wie oft das Ereignis, dessen Wahrscheinlichkeit es einzuschätzen gilt, a priori überhaupt zu erwarten ist.<sup>5</sup> Die Individuen tendieren vielmehr dazu, die Diagnostizität von spezifischen Fall-Informationen gegenüber der Basisraten-Information zu überschätzen. Allerdings konnte festgestellt werden, dass diese systematischen Fehleinschätzungen (*biases*) weniger stark ausgeprägt waren, wenn im

---

<sup>1</sup> Den Grund dafür, dass multiple, konjunktive Ereignisse für wahrscheinlicher als einfache Ereignisse gehalten werden, sehen EINHORN/HOGARTH (1986), S. 3-19 in der Tendenz von Individuen, nach hinreichenden statt nach notwendigen Begründungen zu suchen. Eine hinreichende Begründung für einen Autounfall wäre etwa, dass der Fahrer erstens übermüdet war, zweitens die Kreuzung schwer einblickbar war und drittens Aquaplaning bestand. Als notwendige Bedingung kann sie jedoch keine höhere Wahrscheinlichkeit aufweisen als eine der drei Begründungen. Die Tendenz nach hinreichenden Begründungen kann auch als eine Präferenz nach spezifischen Begründungen bezeichnet werden, die der vollen Natur des Ereignisses gerecht werden.

<sup>2</sup> Vgl. TVERSKY/KAHNEMAN (1982), S. 91-96. Im klassischen, auf TVERSKY/KAHNEMAN zurückgehenden „Linda-Beispiel“ wird die Wahrscheinlichkeit, dass die stereotypisch und als Feministin beschriebene Linda heute Bankangestellte und in der Frauenbewegung aktiv ist, höher eingeschätzt wird, als dass sie einfach nur Bankangestellte ist. Die Konjunktion der Ereignisse und die hieraus resultierende geringere Wahrscheinlichkeit sowohl ‚Bankangestellte‘ als auch ‚in der Frauenbewegung aktiv‘ zu sein, wird hier außer Acht gelassen.

<sup>3</sup> Vgl. HO/MAY (1993), S. 81 m.w.N.

<sup>4</sup> Der Basisratenfehler konnte von TVERSKY/KAHNEMAN in einer Studie nachgewiesen werden, bei der die Probanden die Wahrscheinlichkeit einschätzen sollten, dass ein blaues Taxi in einen Unfall mit anschließender Fahrerflucht verwickelt war. Dabei erhielten die Probanden die Information, dass 85% der Taxis in der Stadt grün und 15% blau sind, und dass anhand einer Zeugenaussage festgestellt werden konnte, dass es sich mit einer Wahrscheinlichkeit von 80% um ein blaues Auto handelte. Die meisten Probanden schätzten die Wahrscheinlichkeit, dass es sich um ein blaues Auto handelte auf 80%. Sie konzentrierten sich demnach auf die Wahrscheinlichkeit, dass ein blaues Auto vom Zeugen als solches erkannt wurde und vernachlässigten dabei, dass die Wahrscheinlichkeit einer irrtümlichen Klassifizierung ebenfalls relevant ist. Vgl. TVERSKY/KAHNEMAN (1982), S. 156-158.

<sup>5</sup> Vgl. JUNGERMANN/PFISTER/FISCHER (1998), S. 167.

Rahmen der Aufgabenstellung explizit auf die Relevanz statistischer Merkmale hingewiesen wurde oder die Basisraten-Information verstärkt hervortrat.<sup>1</sup>

Als *Anker-Heuristik* bezeichnet man die starke Ausrichtung an einer Reizvorlage (Anker) beim Einschätzen von Wahrscheinlichkeiten. Diese Heuristik wird eingesetzt, wenn die Reizvorlage einen expliziten oder impliziten Hinweis für eine erste Approximation bietet. Ausgehend von diesem Anker wird der Schätzwert vom Individuum nach oben oder nach unten adjustiert, wobei die Adjustierung zwar häufig in die richtige Richtung erfolgt, das Ausmaß jedoch zu gering ausfällt.<sup>2</sup> KAHNEMAN/TVERSKY zählen diese Vorgehensweise zu den verkürzten Strategien, mit denen Individuen Urteile abgeben, wenn keine vollständigen Informationen zur Verfügung stehen. Die Anker-Heuristik kann rational sein, sofern der Anker einen Informationsgehalt besitzt. Allerdings konnten KAHNEMAN/TVERSKY nachweisen, dass Individuen sogar zufällig generierte Reizvorlagen als Anker verwenden.<sup>3</sup>

Insgesamt konnte die Anker-Heuristik in zahlreichen Bereichen festgestellt werden und hat sich als ein sehr robustes Phänomen erwiesen.<sup>4</sup> Allerdings ist festzuhalten, dass diese subjektive Fehleinschätzung (*bias*) weniger stark ausgeprägt ist, wenn die Individuen über das für die Wahrscheinlichkeitseinschätzung relevante Faktenwissen verfügten.<sup>5</sup> In der empirischen Prüfungsforschung wurde die Anker-Heuristik häufig im Zusammenhang mit dem Einfluss der Kenntnis ungeprüfter Buchwerte im Rahmen von analytischen Prüfungshandlungen untersucht.<sup>6</sup>

### 3.2.2.2.3 Würdigung der Urteilsbildungsheuristiken

Obwohl dieser Ansatz in der empirischen Prüfungsforschung auf eine große Resonanz gestoßen ist, ist die Erklärung der prüferischen Urteilsbildung durch Heuristiken nicht unumstritten.<sup>7</sup> So ist etwa denkbar, dass ein und dieselbe Heuristik als Erklärung für un-

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 166 m.w.N.

<sup>2</sup> Vgl. JUNGERMANN/PFISTER/FISCHER (1998), S. 171.

<sup>3</sup> TVERSKY/KAHNEMAN befragten in einem Experiment Probanden nach der Anzahl afrikanischer Staaten in der UNO. Die Versuchspersonen beobachteten dabei, wie der Versuchsleiter ein Glücksrad mit Zahlen zwischen 1 und 100 drehte. In Abhängigkeit vom Ergebnis sollten die Versuchspersonen angeben, ob ihre Schätzung größer oder kleiner sei und zusätzlich eine konkrete Zahl benennen. Dabei zeigte sich ein deutlicher Ankereffekt in Richtung der Zufallszahl. Vgl. TVERSKY/KAHNEMAN (1974), S. 1128.

<sup>4</sup> Vgl. SCHREIBER (2000), S. 163 m.w.N.

<sup>5</sup> Vgl. HUSSY (1998), S. 140.

<sup>6</sup> Vgl. hierzu insbesondere Kap. 3.5.2.2.2.

<sup>7</sup> Vgl. im Folgenden LIBBY (1981), S. 69.

terschiedliche Verhaltensmuster herangezogen werden kann. Dieses kann darauf zurückgeführt werden, dass Wahrscheinlichkeitsschlüsse davon abhängig sind, welche Merkmale der jeweiligen Situation dem Entscheidungsverhalten zugrunde gelegt werden. Es besteht also die Gefahr, dass Untersuchungsergebnisse unterschiedlich interpretiert werden können, da nicht eindeutig bestimmbar ist, welche Heuristik eingesetzt wurde.

Darüber hinaus wird kritisiert, dass durch Kombination von Heuristiken quasi jedes mögliche Entscheidungsverhalten erklärt werden kann.<sup>1</sup> Beispielsweise kann sowohl eine konservative als auch eine radikale Modifikation der subjektiven Wahrscheinlichkeitsurteile durch Kombination der Verankerungsheuristik und Repräsentativitätsheuristik erklärt werden. Folglich verliert die Erklärung des Entscheidungsverhaltens an Informationsgehalt und wird tautologisch.

In Anbetracht der aufgeführten Kritikpunkte ist es durchaus berechtigt, die Möglichkeit einer vollständigen Erklärung komplexer Entscheidungssituationen durch einfache Heuristiken anzuzweifeln. Die alleinige Untersuchung von Wahrscheinlichkeitsurteilen reicht sicherlich nicht aus, um das prüferische Vorgehen bei unstrukturierten Prüfungsaufgaben vollends zu erfassen.<sup>2</sup> Hierzu scheint das im folgenden Kap. darzustellende Informationsverarbeitungsmodell, welches sich im Übrigen vollständig von der kritisierten „black-box“-Betrachtung löst, eher geeignet zu sein. Dennoch sollten Heuristiken zur Erforschung der prüferischen Urteilsbildung nicht vollends ad acta gelegt werden. Wenn auch nicht das gesamte Entscheidungsverhalten durch den Einsatz von Heuristiken erklärt werden kann, wäre es unzweckmäßig davon auszugehen, dass nicht einzelne Teilprozesse der menschlichen Informationsverarbeitung durch ihre Verwendung gekennzeichnet sind.

---

<sup>1</sup> Vgl. GANS (1986), S. 190.

<sup>2</sup> Vgl. etwa RUHNKE (2000), S. 240 m.w.N.

### 3.2.2.3 Problemlösungsmodelle

#### 3.2.2.3.1 Modelltheoretische Vorüberlegungen

##### 3.2.2.3.1.1 Der kognitionspsychologische Informationsverarbeitungsansatz als Ausgangspunkt der Betrachtung

Möchte man nicht nur das Ergebnis der Prozesse im Sinne einer „black-box“-Analyse betrachten, sondern auch deren Entstehung verstehen und analysieren, so ist es unumgänglich, die einzelnen mentalen Prozesse im Rahmen der prüferischen Urteilsbildung in den Mittelpunkt der Betrachtung zu rücken.<sup>1</sup> Hierzu erscheint es zweckmäßig, die Erkenntnisse der Kognitionspsychologie heranzuziehen.<sup>2</sup>

Eine Grundlage für die Systematisierung und Integration geistig kognitiver Prozesse bildet der auf NEWELL/SIMON zurückgehende Informationsverarbeitungsansatz.<sup>3, 4</sup> Dieser Ansatz, der mitunter auch als *Problemlösungsansatz* bezeichnet wird, beschreibt kognitive Prozesse als die Verarbeitung von Informationen und betrachtet den Menschen als ein informationsverarbeitendes System.<sup>5</sup> Ein wesentliches Merkmal des Informationsverarbeitungsansatzes ist das Prinzip der begrenzten Rationalität (sog. *bounded rationality*<sup>6</sup>), wonach die menschliche Informationsverarbeitungskapazität gegenüber komplexen Sachverhalten beschränkt ist.<sup>7</sup> Demzufolge ist der Mensch zunächst nicht in der Lage, die komplexe Situation vollständig zu erfassen. Er wird sich vielmehr erst im Laufe der Problembewältigung ein vollständiges bzw. vollständigeres Bild von dem Sachverhalt bilden.<sup>8</sup> Da auch die Jahresabschlussprüfung überwiegend aus Teilbereichen besteht, bei denen sich der Prüfer einer hohen Komplexität gegenüber sieht, scheint der Informationsverarbeitungsansatz zur Abbildung von Prüfungsprozessen besonders geeignet zu sein.<sup>1</sup> Hierbei spielt insbesondere die Herausbildung von prüfungsbezogenen Problemlösungsmodellen eine besondere Rolle, welche die prüferische Vor-

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 14, 81.

<sup>2</sup> Aufgabe dieser Teildisziplin der Psychologie ist es, „elementare Informationsverarbeitungsmechanismen zu identifizieren (...) und schrittweise komplexere Verhaltensebenen durch solche Informationsverarbeitungsmechanismen zu rekonstruieren.“ MACK (1996), S. 98.

<sup>3</sup> Vgl. NEWELL/SIMON (1972), S. 787-868.

<sup>4</sup> Im angloamerikanischen Sprachraum hat sich unter dem Begriff „human information processing in accounting“ eine weitgehend eigenständige Prüfungsforschungsrichtung etabliert, die sich primär am kognitionspsychologischen Informationsverarbeitungsansatz orientiert. Vgl. GEMÜNDEN (1993), S. 844.

<sup>5</sup> Vgl. SCHREIBER (2000), S. 81.

<sup>6</sup> Siehe hierzu ausführlich Kap. 3.2.2.2.2.1.

<sup>7</sup> Vgl. GANS (1986), S. 195.

<sup>8</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 395.

gehensweise bei der Beurteilung von prüfungsbezogenen Sachverhalten möglichst realitätsnah rekonstruieren soll.<sup>2</sup>

### 3.2.2.3.1.2 Skizzierung eines Problemlösungsmodells für die Jahresabschlussprüfung

Aus dem Blickwinkel der Informationsverarbeitung lässt sich die Problemlösung als ein zweistufiger Prozess charakterisieren, der aus dem *Herstellen eines Problemraums* (1.Stufe) und der *Suche im Problemraum* (2.Stufe), und damit dem Finden der eigentlichen Lösung, gekennzeichnet ist.<sup>3</sup> Ein Individuum, das vor der Bearbeitung eines Problems steht, macht sich mit anderen Worten demnach zunächst klar, wie der Ausgangszustand ist und überlegt daraufhin, welche Lösungsmöglichkeiten es zur Verfügung hat, um zu einer erfolgreichen Lösung zu gelangen.

Um einen angemessenen Problemraum aufbauen zu können, muss zunächst gewährleistet sein, dass der Problemlöser den betrachteten Sachverhalt als ein *Problem* erkennt. Hierzu ist es notwendig, dass der Problemlöser durch Wahrnehmung eine *interne Repräsentation* bzw. ein *internes Modell* des extern präsentierten Problemsachverhalts aufbaut, welches ihm ermöglicht, relevante Aspekte des Problems *zu verstehen* und zu erklären.<sup>4, 5</sup> Eine erfolgreiche Problemlösung verlangt demnach zunächst die Bildung

---

<sup>1</sup> So auch SCHREIBER (2000), S. 82; MARTEN/QUICK/RUHNKE (2007), S. 50-51 sowie RUHNKE (2000), S. 290 und MARTEN/QUICK/RUHNKE (2006), S. 393.

<sup>2</sup> Vgl. auch RUHNKE (2006), S. 653.

<sup>3</sup> Vgl. ARBINGER (1997), S. 31-32 sowie MARTEN/QUICK/RUHNKE (2007), S. 50-51 speziell für den Prüfungsbereich.

<sup>4</sup> Vgl. ARBINGER (1997), S. 26. Ähnlich BEDARD/CHI (1993), S. 26 und KOONCE (1993), S. 58-61.

<sup>5</sup> Dass die umfassende Literatur zum Repräsentationsbegriff weitaus mehr unter einer internen Repräsentation subsumiert als die hier eingenommene Sichtweise, lässt sich beispielhaft anhand der Mehrebenenbetrachtung nach OPWIS/LÜER (1996), S. 343-346 veranschaulichen. Demnach handelt es sich bei der internen (Re-)Konstruktion einer äußeren Gegebenheit sowie dem mentalen Operieren lediglich um die erste Ebene der Betrachtung, welche als sog. *psychologische Erfahrungsebene* bezeichnet wird. Die zweite Ebene, die sog. *konzeptuell-fachwissenschaftliche Ebene*, zielt auf die theoretische Rekonstruktion des mentalen Modells eines Individuums ab und bedient sich dabei informellen, theoretisch psychologischen Begriffen. Die weiteren Ebenen begründen sich aus der Verwendung von wissensbasierten Systemen zur computergestützten Rekonstruktion von mentalen Repräsentationen. Bei der *theoretisch-formalen Ebene* werden die in der Regel natürlich-sprachlichen psychologischen Terminologien durch mathematische Strukturen formalisiert. Auf der folgenden *Programmebene* erfolgt die konkrete Implementierung des wissensbasierten Systems im Rahmen einer Programmentwicklung. Die abschließende *Realisierungsebene* beschreibt die Materialisierung des Programms durch die physikalische Installation auf einem bestimmten Rechner. Für die vorliegende Untersuchung sind lediglich die psychologische und die konzeptuell-fachwissenschaftliche Ebene von Bedeutung. Bei der psychologischen Erfahrungsebene geht es zunächst allgemein um die Bestimmung des Begriffs der mentalen Repräsentation eines Problems durch ein Individuum. Dieses ist Gegenstand des vorliegenden Kapitels. Das Hauptanliegen der konzeptuell-fachwissenschaftlichen Ebene ist darin zu sehen, unter Verwendung einer abstrakten, theoretisch psychologischen Begrifflichkeit Modelle der

eines mentalen Surrogats, welches einen Ersatz für das in der äußeren Welt Repräsentierte selbst darstellt.<sup>1</sup> Wichtig ist dabei, dass es sich nicht um eine bloße Abbildung der Realität handelt, sondern um eine (Re-)Konstruktion der Umgebung in Abhängigkeit vom Wissensstand und von der Zielsetzung des Problemlösers.<sup>2</sup> Zu beachten ist hierbei, dass, wie bei jeder Modellbildung, nicht alle Elemente und Relationen des „Originals“ Berücksichtigung finden. Modelle sind immer „Modelle (...) natürlicher oder künstlicher Originale, die ihrerseits auch wieder Modelle sein können (...). Modelle sind ihren Originalen nicht per se zugeordnet, sondern sie erfüllen ihre Funktion stets für bestimmte (...) Akteure und relativ zu bestimmten Zielen und Zwecken.“<sup>3</sup> Bei der internen Repräsentation kann es vorkommen, dass bestimmte Aspekte der Realität nicht enthalten sind oder umgekehrt besonders hervorgehoben werden.<sup>4</sup> Folglich gilt für Problemrepräsentationen, dass sie eine Abstraktion der äußeren Gegebenheiten darstellen, die eine partielle Unvollständigkeit implizieren.

Ist es dem Problemlöser gelungen, ein mentales Modell des Problemsachverhalts aufzubauen, so kann die Problemlösung in die zweite Stufe, dem Absuchen des Problemraums, übergehen. Während das Ziel der ersten Stufe darin liegt, das Problem zu erfassen und zu verstehen, so soll das betrachtete Problem in der zweiten Stufe bearbeitet und gelöst werden. Zur Lösung von Problemen stehen dem Problemlöser unterschiedliche Herangehensweisen zur Verfügung, die man in der Literatur auch als *Problemlösungsstrategien* bzw. *-programme* bezeichnet.<sup>5</sup> Dabei nimmt das in der ersten Stufe aufgebaute interne Modell eine zentrale Rolle ein, da es den Problemlöser erst in die Lage versetzt, bestimmte Handlungen bzw. Konsequenzen von Handlungen mental vorwegzunehmen, ohne dass die Handlungen tatsächlich durchgeführt werden müssen.<sup>6</sup> Man kann dieses auch als die Fähigkeit der „mentalen Simulation realer Prozesse“<sup>7</sup> bezeichnen. Für den Problemlöser birgt dieses den Vorteil, dass er den repräsentierten Sachverhalt adäquat interpretieren kann, indem es ihm ermöglicht wird, die für die Problemlösung relevanten Schritte mental zu durchlaufen und notwendige Schlussfolge-

---

mentalen Repräsentation von Wissen zu rekonstruieren. Hierzu ist das Vorhandensein von Wissen bzw. Wissensstrukturen zwingend notwendig, welche unter Kap. 3.2.2.3.4 eine genauere Darstellung erfahren. Die weiter unten angesiedelten Ebenen sind hingegen dem Bereich der künstlichen Intelligenz zuzuordnen und werden in der vorliegenden Arbeit nicht weiter thematisiert.

<sup>1</sup> Vgl. DAVIS/SHROBE/SZOLOVITS (1993), S. 17.

<sup>2</sup> Ähnlich argumentieren MARTEN/QUICK/RUHNKE (2006), S. 394.

<sup>3</sup> OPWIS (1992), S. 49.

<sup>4</sup> Vgl. OPWIS (1992), S. 13-14. Ähnlich argumentiert auch ARBINGER (1997), S. 78-81.

<sup>5</sup> Vgl. etwa HUSSY (1998), S. 106; BRANDER/KOMPA/PELTZER (1989), S. 136.

<sup>6</sup> Vgl. DAVIS/SHROBE/SZOLOVITS (1993), S. 17.

<sup>7</sup> ARBINGER (1997), S. 26.

rungen zu treffen.<sup>1</sup> Dieses kann als notwendig bezeichnet werden, um schließlich das Problem „extern“ lösen zu können.

Um die vorangegangenen Ausführungen mit Blick auf ein noch zu elaborierendes prüfungsbezogenes Problemlösungsmodell genauer zu spezifizieren, soll in einem nächsten Schritt zum einen erläutert werden, wann ein Sachverhalt als ein „Problem“ deklariert werden kann. Zum anderen soll aufgezeigt werden, welche Vorgehensweisen bzw. Strategien existieren, um ein gegebenes Problem zu lösen (sog. *Problemlösungsstrategien*). Da bereits aufgezeigt wurde, dass die Qualität einer internen Repräsentation und mit ihr auch die Fähigkeit zur Problemlösung im Wesentlichen von den Kenntnissen des Problemlösers abhängen, sollen in einem zweiten Schritt *Wissen* und *Wissensstrukturen* als zentrale Einflussfaktoren der Problemlösungsfähigkeit von Individuen einer ausführlichen Untersuchung unterzogen werden.

### 3.2.2.3.2 Problembegriff

Bemühungen, eine Definition für den Begriff des Problems zu geben, sind in der Literatur zahlreich wieder zu finden.<sup>2</sup> Im Kern gehen alle Definitionen auf die von DUNCKER vorgenommene Gegenstandsbeschreibung zurück.<sup>3</sup> Demnach ist ein Problem als ein unerwünschter *Ausgangszustand* zu beschreiben, der durch eine oder mehrere Handlungen in einen erwünschten *Zielzustand* transformiert werden soll, wobei diese Handlungsalternativen in der Literatur auch als *Operatoren*<sup>4</sup> bezeichnet werden.<sup>5</sup> Allerdings existieren *Barrieren*, die eine Transformation des Ausgangszustands in den Zielzustand verhindern.<sup>6, 7</sup> Als Merkmale eines Problems lassen sich nach diesem Ansatz somit der

---

<sup>1</sup> Vgl. OPWIS (1992), S. 12, 14.

<sup>2</sup> Vgl. etwa DÖRNER (1974), S. 20; DÖRNER (1987), S. 10; DÖRNER/KREUZIG/REITHER/STÄUDEL (1994), S. 302; SMITH (1988), S. 1491; NEWELL/SIMON (1972), S. 72; BROMME/HÖMBERG (1977), S. 76.

<sup>3</sup> Vgl. DUNCKER (1974), S. 1. Demnach handelt es sich um ein *Problem*, „wenn ein Lebewesen ein Ziel hat und nicht ‚weiß‘, wie es dieses Ziel erreichen soll. Wo immer der gegebene Zustand sich nicht durch bloßes Handeln (Ausführen selbstverständlicher Operation) in den erstrebten Zustand überführen lässt, wird das Denken auf den Plan gerufen. Ihm liegt es ob, ein vermittelndes Handeln allerst zu konzipieren.“

<sup>4</sup> In der vorliegenden Arbeit wird hierfür der Begriff der *Problemlösungsstrategien* verwendet, welche den Gegenstand des Kap. 3.2.2.3.3.2 bilden.

<sup>5</sup> Vgl. etwa BRANDER/KOMPA/PELTZER (1989), S. 136.

<sup>6</sup> Vgl. DÖRNER (1987), S. 10.

<sup>7</sup> Barrieren nehmen eine zentrale Rolle für die Abgrenzung eines Problems von einer Aufgabe ein. Zur Bewältigung einer Aufgabe reicht es aufgrund des Fehlens von Barrieren bereits aus, selbstverständliche Operationen auszuführen, um den gewünschten Zielzustand zu erreichen. Da das notwendige Wissen hierzu direkt zur Verfügung steht, ist zur Erreichung eines klar definierten Ziels lediglich der Einsatz bekannter Mittel auf bekannte Weise notwendig. Im Gegensatz zu Problemen können Aufga-

Ausgangszustand, der Zielzustand und die Barrieren aufführen, deren Zusammenhang in Abbildung 5 graphisch festgehalten ist.

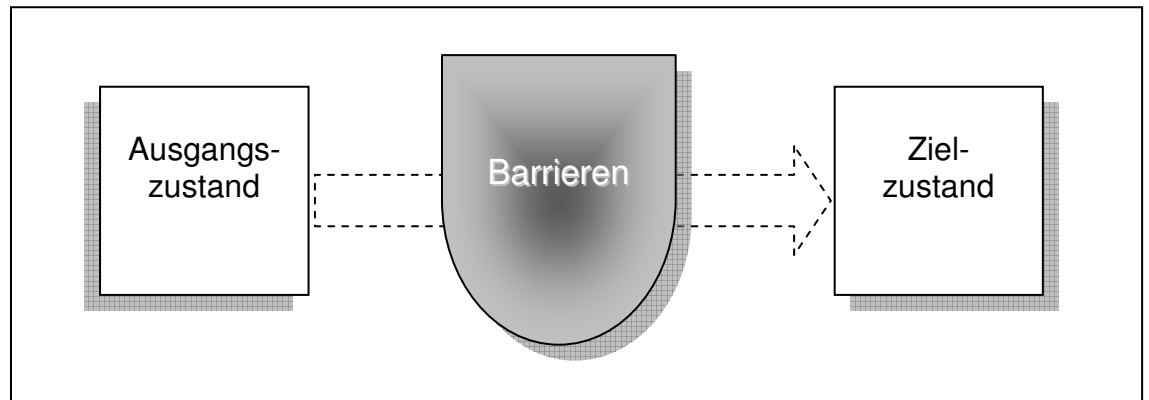


Abbildung 5: Darstellung der Problemelemente

Um eine Ergänzung dieser noch sehr einfachen Definition bemüht, weist GANS darauf hin, dass diese Merkmale lediglich die objektiven Eigenschaften eines Problems betonen. Da es bei Problemen jedoch stets auf die individuellen Eigenschaften eines Subjekts ankommt, erweitert er den Ansatz um eine subjektive und eine zeitliche Dimension. *Subjektive Problemmerkmale* sind das Vorwissen, die Unbestimmtheit, die kognitive Repräsentation und die Motivation.<sup>1</sup> Das *Vorwissen* bezieht sich auf die im Gedächtnis gespeicherten Informationen des Individuums. *Unbestimmtheit*<sup>2</sup> berücksichtigt, dass bestimmte Informationen, die eine unmittelbare Transformation des Ausgangs- in den Zielzustand erlauben, nicht zur Verfügung stehen. Die *kognitive Repräsentation* zielt auf das Bewusstsein der Problemkomponenten durch das Individuum ab. Die *Motivation* letztendlich bedeutet, dass der Unterschied zwischen dem Ausgangs- und Zielzustand ein Spannungsfeld beim Individuum auslöst, das es zum Einsatz von Operatoren bewegt. Durch die zusätzliche Berücksichtigung von subjektiven Problemmerkmalen wird zum Ausdruck gebracht, dass ein Problem ohne die Betrachtung eines Individuums nicht existieren kann.<sup>3</sup> Die *zeitliche Dimension* hingegen gewinnt besonders dann an

---

ben demzufolge „routinemäßig“ ausgeführt werden. Das Lösen eines Problems hingegen bedarf mehr als dem bloßen Abrufen von Wissen, da hier Informationen aus der Umwelt und aus dem Gedächtnis zusammengeführt und verarbeitet werden müssen. Vgl. BRANDER/KOMPA/PELTZER (1989), S. 111-112.

<sup>1</sup> Vgl. GANS (1986), S. 301. GANS greift dabei weitestgehend auf die von BROMME/HÖMBERG (1977), S. 77-78 proklamierten „konstitutiven Merkmale“ zurück und erlangt durch eine Unterteilung in subjektive und objektive Problemmerkmale zusätzlichen Erkenntniswert.

<sup>2</sup> Von GANS auch als Wissenslücke bezeichnet. Vgl. GANS (1986), S. 301.

<sup>3</sup> GANS macht richtigerweise darauf aufmerksam, dass die subjektiven und objektiven Problemelemente nicht unabhängig voneinander sind, da sie sich inhaltlich aufeinander beziehen. Vgl. GANS (1986), S. 301-302.



Bedeutung, wenn für die Transformation des Ausgangszustands in den Endzustand zeitliche Restriktionen zu berücksichtigen sind.<sup>1</sup>

### **3.2.2.3.3 Problemlösungsstrategien**

#### **3.2.2.3.3.1 Genereller Ablauf einer Problemlösung**

Ist die interne Repräsentation des Problemsachverhalts erfolgreich abgeschlossen, so nimmt die Anwendung von Problemlösungsstrategien einen zentralen Bestandteil für das anschließende Absuchen des Problemraums ein. Dabei stehen dem Individuum diverse Strategien zur Verfügung, die teilweise sehr allgemein und im Extremfall nur auf ein einziges Problem bezogen werden können.<sup>2</sup> Der Suchraum beinhaltet theoretisch alle denkbaren Schritte, die sich zwischen dem Ausgangszustand und dem Zielzustand befinden. Die Bearbeitung des Problems muss sich hier jedoch keinesfalls als linearer Prozess darstellen, sondern ist möglicherweise durch verschiedene Rücksprünge und Schleifen gekennzeichnet, wobei auch Kombinationen von Problemlösungsstrategien eingesetzt werden können. Oft führen auch mehrere Wege zur Lösung eines Problems. Auch ist denkbar, dass sich die richtigen Strategien zur Lösung eines Problems überhaupt nicht im Problemraum der jeweiligen Person befinden und diese daher an der Lösung des Problems scheitert.<sup>3</sup>

Welche Problemlösungsstrategie für die Bearbeitung eines Problems besonders geeignet erscheint, hängt einerseits von der jeweiligen Problemart und andererseits von dem jeweiligen Wissen sowie den Fähigkeiten des Individuums ab. Generell lassen sich Problemlösungsstrategien jedoch in einfachere und anspruchsvollere differenzieren. Mitunter werden auch die Heuristiken nach KAHNEMAN/TVERSKY<sup>4</sup> den Problemlösungsstrategien zugeordnet.<sup>5</sup>

---

<sup>1</sup> Vgl. GANS (1986), S. 303.

<sup>2</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 124-125.

<sup>3</sup> Vgl. ARBINGER (1997), S. 33-34.

<sup>4</sup> Siehe hierzu ausführlich Kap. 3.2.2.2.2.

<sup>5</sup> Vgl. etwa MARTEN/QUICK/RUHNKE (2006), S. 394.

### **3.2.2.3.3.2     Arten von Problemlösungsstrategien**

#### **3.2.2.3.3.2.1     Versuch-Irrtum-Analyse**

Die *Versuch-Irrtum-Analyse* kann als die trivialste Problemlösungsstrategie bezeichnet werden, da es sich hier um das bloße „Ausprobieren“ von Handlungsalternativen (Operatoren) handelt.<sup>1</sup> Dabei wird das Individuum (zunächst) zufällig Schritte einleiten und abschätzen, ob sich der gewünschte Erfolg eingestellt hat oder nicht. Misserfolge gehen in das Wissen über die Situation über und schränken somit die Anzahl potenzieller Operatoren ein, so dass das zufällige Probieren in eine systematische Vorgehensweise übergeht.<sup>2</sup>

Obwohl es sich hier um eine sehr simple Strategie handelt, sollte ihr Nutzen nicht unterschätzt werden. Häufig wird sie in sehr frühen Phasen des Bearbeitungsprozesses eingesetzt, um sich eine erste Orientierung zu verschaffen, sich mit dem Problem vertraut zu machen und unterschiedliche Operatoren und deren Auswirkungen auszuprobieren.<sup>3</sup> Demnach ist ihr Einsatz bei allen Problemarten denkbar, in erster Linie jedoch bei Individuen, deren subjektive Merkmale durch ein erhebliches Wissensdefizit über den Bereich, welchem das Problem zuzuordnen ist, gekennzeichnet ist. Da jedoch im Prüfungsbereich einerseits davon auszugehen ist, dass die Prüfer ausbildungsbedingt über ein bestimmtes Maß an notwendigem Wissen verfügen, und andererseits der im Prüfungsbereich mitunter bestehende enorme Fertigstellungs- und Zeitdruck die Anwendung zeitintensiver Vorgehensweisen, denen die *trial and error*-Methode sicherlich zuzurechnen ist, nicht erlaubt, ist diese Problemlösungsstrategie für den Prüfungsbereich sicherlich weniger relevant.

#### **3.2.2.3.3.2.2     Rückwärtsgerichtete Suchstrategien**

##### **3.2.2.3.3.2.2.1     Mittel-Ziel-Analyse**

Die *Mittel-Ziel-Analyse* spielt vor allem bei der Computersimulation menschlichen Problemlösens eine zentrale Rolle, wie es mit Hilfe des von ERNST/NEWELL vorgestell-

---

<sup>1</sup> Die Versuch-Irrtum-Analyse wird in Anlehnung an die angloamerikanische Literatur mitunter auch als *trial and error-strategy* bezeichnet. Vgl. ARBINGER (1997), S. 51.

<sup>2</sup> Als bekanntestes Beispiel kann wohl auf das Memory-Spiel verwiesen werden, bei dem versucht werden muss, jeweils zwei Karten mit dem gleichen Spiel aufzudecken. Vgl. BRANDER/KOMPA/PELTZER (1989), S. 126.

<sup>3</sup> Vgl. ARBINGER (1997), S. 51-53.

ten Computerprogramms „General Problem Solvers“ ausführlich untersucht wurde.<sup>1</sup> Ausgangslage ist eine genaue Beschreibung des Ausgangs- und Zielzustands und der verfügbaren Operatoren. Die unterstellte Vorgehensweise lässt sich dabei als eine systematische Abfolge der nachfolgend beschriebenen Schritte darstellen.<sup>2</sup> Zunächst werden die Unterschiede zwischen Ausgangs- und Zielzustand festgestellt (1. Schritt). Daraufhin werden die notwendigen Operatoren festgelegt (2. Schritt) und ausgeführt (3. Schritt). Abschließend wird eine erneute Bestimmung der Unterschiede zwischen dem aktuellen und dem Zielzustand vorgenommen (4. Schritt).

Charakteristisch für die *Mittel-Ziel-Analyse* ist ihre rückwärtsorientierte Suchrichtung. Ausgehend von dem erwünschten Zielzustand zeichnet sie sich „durch die rekursive Konstruktion von Zwischenzielen mit wiederholter Durchführung von Rückwärtssuche bis zum Erreichen des Ausgangszustands aus.“<sup>3</sup> Die Wiederholung der Schritte wird so lange ausgeführt, bis der Ausgangszustand vollkommen mit dem Zielzustand übereinstimmt.<sup>4</sup> So wird der Ausgangszustand schrittweise in den gewünschten Zielzustand überführt und somit das Problem gelöst.<sup>5</sup>

Allerdings stellt sich hier die Frage, inwiefern ein Computerprogramm in der Lage ist, menschliches Problemlösen zu simulieren. NEWELL/SIMON konnten in empirischen Untersuchungen nachweisen, dass sich die Vorgehensweisen von Individuen bei der Bearbeitung von Interpolationsproblemen im Sinne der Mittel-Ziel-Analyse einordnen lassen.<sup>6</sup> Hier sind allerdings auch schon die Grenzen der Mittel-Ziel-Analyse zu sehen, da sie nur auf einfache Probleme bezogen werden kann. Menschen gehen hingegen nicht so systematisch wie Computer vor und führen nicht immer eine genaue Analyse der Situation durch.<sup>7</sup> Außerdem planen Menschen eher kleine Handlungssequenzen, während das Computerprogramm nur einzelne Schritte plant und ausführt.<sup>8</sup> Gewisse Mängel in der systematischen Vorgehensweise können Menschen dadurch kompensieren, dass aus großen Mengen gespeicherter Informationen gezielt bestimmte Elemente und Strategien herausgegriffen werden, ohne dabei den ganzen Speicher abzusuchen.<sup>9</sup>

---

<sup>1</sup> Vgl. ERNST/NEWELL (1969), insbesondere S. 35-62.

<sup>2</sup> Vgl. im Folgenden BRANDER/KOMPA/PELTZER (1989), S. 130-131.

<sup>3</sup> GRUBER/MANDL (1996), S. 588.

<sup>4</sup> DÖRNER (1987), S. 74-75 bezeichnet diese Methode auch als die „Reduktionsmethode“.

<sup>5</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 130-131.

<sup>6</sup> Vgl. NEWELL/SIMON (1972), S. 707-747.

<sup>7</sup> Vgl. GREENO (1978), S. 247.

<sup>8</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 131.

<sup>9</sup> BRANDER/KOMPA/PELTZER (1989), S. 131.

Zusammenfassend lässt sich sagen, dass sich die Mittel-Ziel-Analyse auf die Bearbeitung von Problemen einschränken lässt, bei denen klare Vorstellungen über den Ausgangs- und Zielzustand vorliegen und Operatoren sowie deren Auswirkungen bekannt sind.

### **3.2.2.3.3.2.2.2 Generate and test-Strategie**

Eine weitere rückwärtsgerichtete Suchstrategie ist die *generate and test-Strategie*.<sup>1</sup> Im Gegensatz zur Mittel-Ziel-Analyse sind hier allerdings keine klaren Vorstellungen über die Problemmerkmale notwendig, so dass ihr eine höhere Eignung zur Beschreibung der menschlichen Vorgehensweise bei der Bearbeitung von Problemen zu bescheinigen ist.

Die *generate and test-Strategie* ist allgemein dadurch gekennzeichnet, dass von einem möglichen Zielzustand aus versucht wird, Zustände zu bilden, von denen aus der erwünschte Zielzustand erreicht werden soll.<sup>2</sup> D.h. es wird vom Zielzustand aus zurückgearbeitet und die so gefundenen Zustände werden als Zwischenziele definiert.<sup>3</sup> Bezugnehmend auf das in dieser Arbeit zugrunde gelegte Modell bedeutet diese rückwärtsgerichtete Suchrichtung, dass zunächst eine Hypothese generiert wird und anschließend die diese Hypothese bestätigenden Fakten gesucht werden.<sup>4</sup> Können die beschafften Informationen die generierte Hypothese nicht verifizieren, so wird diese verworfen und es wird eine neue Hypothese generiert. Die *generate and test-Strategie* zeichnet sich daher durch Wiederholungen und Rekursivität aus und wird mitunter auch als „wandering about“ umschrieben.<sup>5</sup> In der Literatur wird diese Vorgehensweise auch als die Novizenstrategie bezeichnet, da an sie keine klare Erfassung der objektiven Problemmerkmale vorausgesetzt wird.<sup>6</sup>

### **3.2.2.3.3.2.3 Vorwärtsgerichtete Suchstrategie**

Von einer *vorwärtsgerichteten Suchstrategie* ist immer dann die Rede, wenn sich ein Individuum vom Ausgangszustand aus Anhaltspunkte im Sachverhalt sucht, um von

---

<sup>1</sup> Vgl. etwa ROTHE/SCHINDLER (1996), S. 43 oder BEDARD/CHI (1993), S. 27-28.

<sup>2</sup> Vgl. GRUBER/MANDL (1996), S. 588.

<sup>3</sup> Vgl. ARBINGER (1997), S. 63-64.

<sup>4</sup> Vgl. auch ROTHE/SCHINDLER (1996), S. 43 oder BEDARD/CHI (1993), S. 27.

<sup>5</sup> Vgl. SCHREIBER (2000), S. 226.

<sup>6</sup> Vgl. etwa ROTHE/SCHINDLER (1996), S. 42-43; BEDARD/CHI (1993), S. 27-28.

dort aus Fortschritte in Richtung des erwünschten Zielzustands durchzuführen.<sup>1</sup> Diese klare zielgerichtete Vorgehensweise ermöglicht es dem Individuum, eine Beschränkung auf die für die Problemlösung notwendigen Handlungen vorzunehmen und sich weniger durch irrelevante Gegebenheiten der jeweiligen Situation beeinflussen zu lassen.<sup>2</sup> Für das in dieser Arbeit unterstellte Modell bedeutet eine vorwärtsgerichtete Suchrichtung, dass von den gegebenen Fakten aus Hypothesen generiert werden.<sup>3</sup> Im Gegensatz zur rückwärtsgerichteten Suchstrategie konzentriert man sich bei der vorwärtsgerichteten Suchstrategie von vornherein auf die wichtigen Aspekte und Informationen des Problemsachverhalts, was eine effizientere und effektivere Bearbeitung des Problems gewährleistet.<sup>4</sup>

Allerdings ist für eine vorwärtsgerichtete Suchstrategie der Rückgriff auf vorhandene Wissensstrukturen notwendig, der eine schnelle Zuordnung des gegebenen Sachverhalts und der notwendigen Operatoren ermöglicht. Daher wird diese Vorgehensweise in der Literatur auch häufig als *Expertenstrategie* bezeichnet.<sup>5</sup> Dieses soll jedoch nicht bedeuten, dass Experten generell eine vorwärtsgerichtete Suchstrategie verwenden. Vielmehr ist davon auszugehen, dass Experten abhängig von den verfügbaren Wissensstrukturen beide Suchrichtungen einsetzen können.<sup>1</sup> Stößt ein Experte auf einen Problemsachverhalt, der einer gespeicherten Wissensstruktur entspricht, so löst dieses die Vornahme einer Handlung aus (etwa die Generierung einer Hypothese), was dann als vorwärtsgerichtet bezeichnet werden kann. Diese vorwärtsgerichtete Problemlösungsstrategie darf allerdings nicht mit der Bearbeitung einer Aufgabe verwechselt werden, da hier weniger von einer routinemäßigen Lösung ausgegangen werden kann, als vielmehr von einem weiterhin kognitiv aufwendigen Deuten und Interpretieren des Problemsachverhalts. Treffen hingegen Experten auf Situationen, die einen Rückgriff auf vorhandene Wissensstrukturen nicht erlauben, so wenden auch sie, ähnlich wie Novizen, eine rückwärtsgerichtete Suchstrategie an.<sup>2</sup>

---

<sup>1</sup> Vgl. GRUBER/MANDL (1996), S. 588.

<sup>2</sup> Vgl. SCHREIBER (2000), S. 225.

<sup>3</sup> Vgl. auch ROTHE/SCHINDLER (1996), S. 43 oder BEDARD/CHI (1993), S. 27.

<sup>4</sup> Vgl. BONNER/PENNINGTON (1991), S. 14.

<sup>5</sup> Vgl. etwa GRUBER/MANDL (1996), S. 588; SCHREIBER (2000), S. 225-226.

#### 3.2.2.3.3.2.4 Situations- und Zielanalyse

Bei der *Situations- und Zielanalyse* nehmen die systematische Erfassung des Ausgangs- und Zielzustands und die Kenntnis über die notwendigen Operatoren sowie deren Auswirkungen eine zentrale Rolle für die Problemlösung ein. Damit weist sie eine starke Ähnlichkeit zu der weiter oben dargestellten Mittel-Ziel-Analyse auf.<sup>3</sup> Allerdings zielt die Situations- und Zielanalyse weniger auf eine *Computer-Mensch-Metapher* ab, da zur Problembearbeitung zunächst keine vollkommene Klarheit über den Ausgangs- und Zielzustand sowie die notwendigen Operatoren vorliegen muss. Vielmehr erfolgt eine Präzisierung der objektiven Problemmerkmale erst im Laufe der Problemlösung. Die Situations- und Zielanalyse weist gegenüber der Mittel-Ziel-Analyse somit den Vorteil auf, dass sie eher für das menschliche Problemlösen herangezogen werden kann.

Die *Situations- und Zielanalyse* beginnt, soweit möglich, mit der Herausarbeitung der spezifischen Elemente des Ausgangs- und des Zielzustands eines Problems. Anschließend kann das Individuum feststellen, welche Gemeinsamkeiten und Unterschiede vorliegen und welche Operatoren geeignet sind, um die gewünschte Transformation zu erreichen (*Operatorenanalyse*).<sup>4</sup> Die Betonung der systematischen Vorgehensweise soll dabei nicht bedeuten, dass die Situations- und Zielanalyse immer chronologisch mit der Erfassung des Ausgangszustands beginnen muss. Vielmehr ist in Abhängigkeit der individuellen Eigenschaften auch eine umgekehrte Reihenfolge denkbar. Damit kann die Situations- und Zielanalyse sowohl mit einer rückwärts- als auch mit einer vorwärtsorientierten Suchrichtung kombiniert werden.

Die *Situationsanalyse* kann in eine *Konfliktanalyse* und eine *Materialanalyse* unterteilt werden.<sup>5</sup> Im Rahmen der Konfliktanalyse ist festzustellen, welche Elemente in einer Ausgangssituation unerwünscht sind und ob diese verändert werden können. Die Materialanalyse hingegen untersucht den Ausgangszustand darauf hin, welche Elemente in den Zielzustand übernommen werden können bzw. welche Elemente zur Konstruktion einer Lösung herangezogen werden können. Im Anschluss an die Situationsanalyse erfolgt die Analyse des erwünschten Zielzustands. Hier „werden Prioritäten für die Lö-

---

<sup>1</sup> Vgl. hierzu ausführlich O'DONNELL (2004), S. 75-76 und O'DONNELL/LEHMAN (1999), S. 245-247. O'DONNELL/LEHMAN bezeichnen die vorwärtsgerichtete Suchstrategie als „configural processing“ und die rückwärtsgerichtete Suchstrategie als „serial processing“.

<sup>2</sup> Vgl. BEDARD/CHI (1993), S. 27.

<sup>3</sup> Vgl. hierzu Kap. 3.2.2.3.3.2.2.1.

<sup>4</sup> Vgl. DÖRNER (1987), S. 60.

<sup>5</sup> Vgl. im Folgenden BRANDER/KOMPA/PELTZER (1989), S. 127.

sung und Kriterien für den angestrebten Zielzustand herausgearbeitet“.<sup>1</sup> Mit dem anschließenden Vergleich beginnt die Phase der *Operatorenanalyse*, in der die möglichen Operatoren auf ihren Beitrag zur Transformation des Ausgangs- in den Zielzustand untersucht werden.<sup>2</sup>

Für die Anwendung der Situations- und Zielanalyse reicht das bloße Erfassen der Problemelemente nicht aus. Vielmehr ist das Vorhandensein von Wissen notwendig, da Schlussfolgerungen über das vorhandene Material getätigt und die Wirkungsweisen der Operatoren abgeschätzt werden müssen.<sup>3</sup>

Der Einsatz dieser Problemlösungsstrategie scheint insbesondere für solche Probleme sinnvoll, bei denen der Ausgangszustand einerseits oder der Zielzustand andererseits nicht eindeutig definiert sind. In beiden Fällen kann eine allmähliche Präzisierung des fehlenden Zustands durch „das gedankliche Hin- und Herspringen vom Ausgangs- in den Zielzustand“<sup>4</sup> erfolgen. Insgesamt ist der Situations-Ziel-Analyse eine bedeutendere Rolle beizumessen als der Mittel-Zweck-Analyse.

### **3.2.2.3.3.2.5 Umstrukturierung und Neuentdeckung**

#### **3.2.2.3.3.2.5.1 Einstellungs- und Fixierungseffekte**

Ein häufiger Grund, weshalb Individuen an der Lösung eines Problems scheitern, liegt darin, dass sie sich in einem Problemraum befinden, der kleiner als der Raum ist, in dem sich die Lösung des Problems tatsächlich befindet.<sup>5</sup> Die Ursache einer solchen Fehleinschätzung des Sachverhalts ist in dem Auftreten von Einstellungs- und Fixierungseffekten zu sehen. Derartige einer Problemlösung im Wege stehende Phänomene treten häufig aufgrund von Denkgewohnheiten auf.<sup>6</sup>

*Einstellungseffekte*<sup>7</sup> führen dazu, dass man aufgrund früherer Erfahrungen zur Anwendung bestimmter Operatoren tendiert.<sup>1</sup> Andere Operatoren werden für ein neues Prob-

---

<sup>1</sup> BRANDER/KOMPA/PELTZER (1989), S. 127.

<sup>2</sup> Vgl. DÖRNER (1987), 61.

<sup>3</sup> Vgl. DÖRNER (1987), 61.

<sup>4</sup> BRANDER/KOMPA/PELTZER (1989), S. 128.

<sup>5</sup> Vgl. DÖRNER (1987), S. 77.

<sup>6</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 134.

<sup>7</sup> Dieser so genannte Einstellungseffekt zeigt sich zum Beispiel im so genannten Neun-Punkte-Problem, bei dem neun Punkte mit vier Strichen ohne Absetzen verbunden werden sollen. Da eventuell viele im Geometrieunterricht gelernt haben, Punkte zu verbinden, kommen deswegen vielleicht wenige Menschen auf die Idee, dass für die richtige Lösung die Striche über die Punkte hinaus gezogen werden müssen. Vgl. ARBINGER (1997), S. 71 sowie ANDERSON (2001), S. 269.

lem nicht mehr in Erwägung gezogen, obwohl sie dem Individuum prinzipiell zur Verfügung stehen und in der aktuellen Situation besser dazu geeignet sind, das vorliegende Problem zu lösen.<sup>2</sup> Da die bevorzugten Operatoren jedoch häufig zum Erfolg geführt haben, vertraut das Individuum auf ihrer erneuten Anwendung und erkennt, dass die vorliegende Situation den Einsatz anderer Operatoren erfordert. „Einstellungseffekte resultieren aus einer Verstärkung des Wissens, das für die Lösung eines bestimmten Problemtyps relevant ist.“<sup>3</sup> Wissen kann demnach auch als Hindernis betrachtet werden, das die erfolgreiche Bearbeitung eines Problems behindert.<sup>4</sup>

*Fixierungseffekte*<sup>5</sup> liegen immer dann vor, wenn Individuen nicht in der Lage sind, Problemmerkmale aus ihrer gewohnten Umwelt herauszulösen und auf etwas anderes zu beziehen.<sup>6</sup> Fixierungen sind darauf zurückzuführen, dass das Individuum darauf festgelegt ist, einem bestimmten Objekt eine bestimmte Funktion zuzuordnen. ANDERSON spricht daher auch von der „funktionalen Fixierung“.<sup>7</sup> Diese starke Verknüpfung eines Objekts an eine bestimmte Funktion ist nicht zwangsläufig als problematisch zu bezeichnen, da sie in vielen Situationen durchaus auch zu einer erfolgreichen Problemlösung führen kann. Kritisch wird es allerdings dann, „wenn die Lösung eines Problems sozusagen die ‚ungebundene‘, von der ‚Norm‘ abweichende Verwendung eines Gegenstands erfordert.“<sup>8</sup>

Haben sich Einstellungs- oder Fixierungseffekte bei der Erfassung eines Problemsachverhalts ergeben, so hängt der Erfolg einer Problemlösung maßgeblich von der Fähigkeit des Individuums ab, Strategien zu entwickeln, die es ihm ermöglichen, diese Beeinträchtigungen zu überwinden. Derartige Strategien sollten auf die Neuentdeckung von Operatoren oder auf die Umstrukturierung des Sachverhalts ausgerichtet sein.<sup>9</sup> Hierzu

---

<sup>1</sup> Vgl. ANDERSON (2001), S. 268.

<sup>2</sup> Vgl. ARBINGER (1997), S. 71.

<sup>3</sup> Vgl. ANDERSON (2001), S. 271.

<sup>4</sup> Vgl. KREMS (1996), S. 81. SCHREIBER (2000), S. 253 spricht in diesem Zusammenhang von einem Wissens-Flexibilitäts-Paradoxon, was nur durch die Konzeptualisierung von Flexibilität als wissensbasierte, kognitive Fähigkeit gelöst werden kann.

<sup>5</sup> Fixierungseffekte zeigen sich zum Beispiel beim „Zwei-Seile-Problem“, wo eines der Seile mit Hilfe der Zange zu einem Pendel umfunktioniert und in Schwingung gebracht werden muss. Nun kann man das eine Seil in die Hand nehmen und mit der anderen Hand das schwingende Seil ergreifen. Die Schwierigkeit dieser Problemlösung ist darin zusehen, dass man die Zange nicht als Gewicht wahrnimmt, welches sich als Pendel „umfunktionieren“ lässt, sondern in seiner funktionalen Bedeutung als Werkzeug. Vgl. ANDERSON (2001), S. 267.

<sup>6</sup> Vgl. ANDERSON (2001), S. 267. ARBINGER (1997), S. 72-74 verwendet für dieses Phänomen den Begriff „Gebundenheit“.

<sup>7</sup> Vgl. ANDERSON (2001), S. 267.

<sup>8</sup> ARBINGER (1997), S. 73.

<sup>9</sup> Vgl. DÖRNER (1987), S. 79.



kommen zwei Arten von Heuristiken in Betracht, die im Folgenden ausführlich diskutiert werden.

### 3.2.2.3.3.2.5.2 Umstrukturierungsheuristiken

Umstrukturierungsheuristiken können als Strategien bezeichnet werden, die es dem Individuum ermöglichen, den Sachverhalt aus einer anderen Perspektive zu betrachten und ihm auf diese Weise Operatoren bereitstellen, die eigentlich bekannt, aber aufgrund der eingeschränkten Sichtweise nicht in Verbindung mit der Lösung des betrachteten Problems gebracht werden.<sup>1</sup> Umstrukturierungen müssen dabei nicht unbedingt eine völlig neue Herangehensweise an Sachverhalte bedeuten, sondern oft ist nur eine Veränderung der Bedeutung sowie Zuordnung einzelner Objekte der Situation notwendig. Dies setzt aber zumindest eine gewisse Flexibilität im Denken voraus, da meist Übertragungen auf andere Abstraktionsniveaus, Darstellungsformen und Realitätsformen notwendig sind, um das Problem erfolgreich zu bewältigen.<sup>2</sup> Möglichkeiten eines initiierten Anstoßes zur Umstrukturierung eines analysierten Ausgangszustands stellen nach DUNCKER

- das Ausfällen des Gemeinsamen,
- die Resonanzwirkung eines tauglichen Signalelements sowie
- die willkürliche Änderung der Gestaltauffassung

dar.<sup>3</sup>

Beim *Ausfällen von Gemeinsamkeiten* analysiert man alle fehlgeschlagenen Operatoren auf ein ihnen gemeinsames Merkmal und versucht dann eine andere Richtung einzu-

---

<sup>1</sup> Vgl. DÖRNER (1987), S. 80.

<sup>2</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 133-134.

<sup>3</sup> Vgl. im Folgenden DUNCKER (1974), S. 36. Als weitere Möglichkeit führt DUNCKER die Sättigung auf. Hierbei kippt das Problem in eine andere Sichtweise um, wenn der Sachverhalt immer aus der gleichen Perspektive betrachtet wird. Als Beispiel für die durch Sättigung initiierten Umstrukturierungen lassen sich die sog. Kippfiguren aus dem Bereich der optischen Wahrnehmung aufführen. Bei der Betrachtung der „Schröderschen Treppe“ erfolgt die Vertauschung von vorn und hinten. Ob dieses spontane „Umkippen“ durch Sättigung auch bei den in dieser Arbeit betrachteten anspruchsvolleren Problemen angenommen werden kann, ist allerdings mehr als fraglich und soll daher nicht weiter betrachtet werden.

schlagen.<sup>1</sup> Diese Methode wird nach KAPLAN/SIMON auch als die *Suche nach den Invarianten* in einer Situation bezeichnet.<sup>2</sup>

Die *Resonanzwirkung eines tauglichen Signalelements* kann als ein zweistufiger Prozess bezeichnet werden, bei dem zunächst ein abstrakter Suchraum aufgebaut wird, der dem Individuum eine diffuse Vorstellung von den für die Problemlösung notwendigen Operatoren gibt. Anschließend wird der zu bearbeitende Sachverhalt auf Elemente analysiert, die in den aufgebauten Suchraum hineinpassen, d.h. eine Resonanzwirkung hervorrufen.<sup>3</sup> Da die Resonanzwirkung maßgeblich auf das Wiedererkennen von Ähnlichkeiten beruht, spielt die Erfahrung eines Menschen und damit zusammenhängend das Vorhandensein von Wissen eine dominierende Rolle für die erfolgreiche Lösung eines Problems durch Umstrukturierung.<sup>4</sup>

Bei der *willkürlichen Änderung der Gestaltauffassung* wendet das Individuum systematisch unterschiedliche Operatoren an, ohne dabei auf deren subjektiv eingeschätzte Angemessenheit zu achten. Unter Umständen muss sich ein Individuum dazu „zwingen“, eine Operation „durchzudenken“, um festgefahrene Auffassungen über die Situation aufzugeben und gleichzeitig neue Ansätze für die Problemlösung zu erhalten.<sup>5</sup> Diese Vorgehensweise kann auch als systematisches Versuch-Irrtum-Verhalten bezeichnet werden, bei dem Aspekte des Problemsachverhalts, die bisher nicht beachtet wurden, in den Vordergrund rücken und somit Verknüpfungen zwischen bisher nicht zusammenhängenden Merkmalen beleuchten.<sup>1</sup>

### 3.2.2.3.2.5.3 Neuentdeckungsheurismen

Sieht sich ein Individuum einem Sachverhalt gegenüber gestellt, dem es bisher noch nicht begegnet ist, so hängt der Erfolg der weiteren Problemlösung entscheidend von

---

<sup>1</sup> Beim sog. Neun-Punkte-Problem könnte das Individuum auf diese Weise darauf kommen, dass der Grund für seine fehlgeschlagenen Lösungsversuche darin liegt, dass es bei seinen bisherigen Ansätzen innerhalb der vorgegebenen Anordnungen verblieben ist und so auf die umstrukturierungsinitiierende Idee kommt.

<sup>2</sup> Vgl. KAPLAN/SIMON (1990), S. 382, 403-407.

<sup>3</sup> Als Beispiel für eine derartige Vorgehensweise lässt sich die Lösung des sog. *Schachtelproblems* anführen. Man sucht nach einer rechtwinkligen Unterlage, die man an der Wand befestigen kann, und die als Auflage für eine Kerze dient. Die Analyse des Problemraums ergibt, dass eine Streichholzschatel, in der sich die Heftzwecken befinden, diese Funktion erfüllt und somit eine Resonanzwirkung für die Lösung des Problems auslöst.

<sup>4</sup> So auch ARBINGER (1997), S. 75-76.

<sup>5</sup> Beispielhaft lässt sich die Lösung des sog. Zwei-Seil-Problems heranziehen, bei dem das Individuum auch scheinbar unkonventionelle Operatoren, wie etwa die unfunktionale Verwendung einer Zange als Gewicht, in Betracht ziehen muss.

der Fähigkeit des Individuums ab, Strategien zu entwickeln, die ihm die Erweiterung seines Operatoreninventars gestatten.<sup>2</sup> Hierzu sind die Neuentdeckungsheuristiken zu zählen, die in die Analogiebildung einerseits und die Abstraktion andererseits klassifiziert werden können.<sup>3</sup>

Bei der *Analogiebildung* ist die Vorgehensweise eines Individuums dadurch gekennzeichnet, dass es versucht, neue Erkenntnisse über etwas bisher Unverstandenes zu gewinnen, indem es Parallelen zu etwas bereits Bekanntem zieht.<sup>4</sup> Im Gegensatz zur Umstrukturierung ist die Analogiebildung nicht darauf ausgelegt, bekannte, jedoch nicht in Betracht gezogene Operatoren zu reaktivieren, sondern „neue“ Operatoren zu entdecken, die das Individuum zumindest in dem betrachteten Sachverhalt noch nicht angewandt hat. Insofern bietet diese Strategie den Vorteil, dass durch eine einfache Transferleistung die Notwendigkeit entfällt, eine aufwendige Neuentwicklung von für die Lösung des Problems notwendigen Operatoren vorzunehmen.<sup>5</sup> Das wichtigste Merkmal der Analogiebildung liegt darin, dass vorhandenes Wissen über einen bestimmten Sachverhalt auf bisher unbekannte, jedoch strukturähnliche Sachverhalte übertragen wird.<sup>6</sup> In der Literatur werden diese beiden Sachverhalte häufig als Ziel- bzw. Quellproblem bezeichnet, wobei Erstgenanntes das aktuell zu bearbeitende und Letztgenanntes das in der Vergangenheit bereits gelöste Problem darstellt.<sup>7, 8</sup>

Unter *Abstraktion* versteht man „eine Analyse, die das Beiläufige, Unwesentliche, die verfälschenden Umstände eliminiert und auf diese Weise das Wesentliche der Erscheinungen extrahiert.“<sup>9</sup> Von zentraler Bedeutung ist hierbei die Fähigkeit, einen komplexen

---

<sup>1</sup> Vgl. ARBINGER (1997), S. 76.

<sup>2</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 132.

<sup>3</sup> Weitere Neuentdeckungsstrategien stellen etwa die *Modellbildung*, *Metaphorik* und *Imagination* dar. Vgl. etwa BRANDER/KOMPA/PELTZER (1989), S. 132-133 oder ARBINGER (1997), S. 77-90.

<sup>4</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 97.

<sup>5</sup> Vgl. NOVICK (1988), S. 510.

<sup>6</sup> Vgl. ARBINGER (1997), S. 92-93.

<sup>7</sup> Vgl. etwa ANDERSON (2001), 249; NOVICK (1988), S. 510-511 oder HESSE/HAHN (1994), S. 91.

<sup>8</sup> Nach DÖRNER (1987), 82-83 kann die Analogiebildung in die vier Teilschritte Abstraktion, Modellsuche, Rückübertragung und Prüfung eingeteilt werden. Bei der *Abstraktion* muss das Zielproblem zunächst um konkrete Sachverhaltsmerkmale dekontextualisiert werden, da ein Sachverhalt einem anderen nie völlig, sondern nur in Bezug auf die wesentlichen Merkmale entsprechen kann. Die *Modellsuche* zielt auf die Suche nach einem für den Transfer geeigneten Quellproblem ab. Das Quellproblem stellt dabei eine andere Konkretisierung des in der Abstraktionsphase gewonnenen Sachverhalts als das Zielproblem dar. In der darauf folgenden Phase der *Rückübertragung* wird das relevante Wissen über das Quellproblem auf das Zielproblem transferiert. Abschließend erfolgt die *Prüfung*, ob das durch Rückübertragung gewonnene neue Wissen auch tatsächlich auf das Zielproblem übertragen werden kann. Es ist folglich zu erkennen, dass die Analogiebildung untrennbar mit der Abstraktion verbunden ist. Dass Neuentdeckungsstrategien sich nicht einander ausschließen, stellt auch ARBINGER (1997), S. 77, 87 heraus.

<sup>9</sup> RUBINSTEIN (1972), S. 41.

Sachverhalt in seine einzelnen Bestandteile aufzuspalten sowie deren Wichtigkeit, Dringlichkeit und (subjektive) Erfolgswahrscheinlichkeit zu beurteilen.<sup>1, 2</sup> Anschließend werden die wesentlichen Bestandteile neu organisiert, so dass sich der Problemlöser ein verallgemeinertes, kontextunabhängiges Bild über den zu beurteilenden Sachverhalt konstruieren kann.<sup>3</sup> Abstraktion ermöglicht es Individuen demnach, sich weniger durch widersprüchliche oder verwirrende Informationen des gegebenen Kontexts beeinflussen zu lassen, d.h. sie nehmen Umweltreize unabhängig vom betrachteten Sachverhalt wahr, um darauf aufbauend eine analytisch geprägte Problemlösung durchzuführen.<sup>4</sup>

### **3.2.2.3.4 Wissen und Wissensstrukturen**

#### **3.2.2.3.4.1 Der Wissensbegriff und die unterschiedlichen Arten des Wissens**

Grundvoraussetzung einer erfolgreichen Problemlösung bei analytischen Prüfungshandlungen ist zweifellos das Vorhandensein einer umfassenden Wissensbasis. Allgemein versteht man unter Wissen die Gesamtheit der im Gedächtnis gespeicherten Informationen, auf die durch Abruf oder mentale Weiterverarbeitung zurückgegriffen werden kann.<sup>5</sup>

Während bei der Definition des Wissensbegriffs ein weitgehender Konsens in der Literatur besteht, herrscht jedoch eine weitgehende Uneinigkeit darüber, welche unterschiedlichen Arten von Wissen zu differenzieren sind. Begriffspaare wie

- deklaratives vs. prozedurales Wissen,
- Fakten- vs. Handlungswissen,
- semantisches vs. episodisches Wissen oder
- implizites vs. explizites Wissen

stellen nur einige Beispiele aus einer Vielzahl von Klassifizierungen von Wissen dar, wobei die einzelnen Unterteilungen keineswegs als überschneidungsfrei bezeichnet

---

<sup>1</sup> Vgl. PUTZ-OSTERLOH (1983), S. 103.

<sup>2</sup> Da die Aufspaltung einer Gesamtheit in ihre einzelnen Bestandteile einen unmittelbaren Zusammenhang zum Analysebegriff aufweist, wird die individuelle Fähigkeit zur Abstraktion auch häufig als *analytisches Denkvermögen* bezeichnet. Vgl. etwa SCHREIBER (2000), S. 231-237.

<sup>3</sup> Vgl. PINCUS (1990), S. 151-152.

<sup>4</sup> Vgl. MILLS (1996), S. 51.

<sup>5</sup> Vgl. OPWIS (1992), S. 50. Ähnliche Definitionen finden sich bei LIBBY (1995), S. 181; BODE (1997), S. 458-459; ROTHE/SCHINDLER (1996), S. 38 oder RUSSO JR. (1994), S. 102-103.

werden können.<sup>1</sup> In der Prüfungsliteratur wird der Klassifizierung in *deklaratives* vs. *prozedurales Wissen* die größte Bedeutung beigemessen.<sup>2</sup> Daher soll diese Unterscheidung auch die Grundlage für die vorliegende Arbeit bilden.

Deklaratives Wissen ist das Wissen eines Individuums über Begriffe und deren Beziehungen zueinander, über Sachverhalte, Ereignisse, Objekte und Personen und kann durch Mitteilung oder Lesen angeeignet werden.<sup>3</sup> Deklaratives Wissen ist Wissen, dessen wir uns bewusst sind und über das wir zu berichten im Stande sind.<sup>4</sup> Solches Wissen wir häufig auch als *Wissen, dass ...* bezeichnet.<sup>5</sup>

Deklaratives Wissen kann in *episodisches* und *semantisches Wissen* unterteilt werden. Episodisches Wissen stellt Wissen über persönliche Erfahrungen und Erlebnisse (etwa mit bestimmten Prüfungsaufgaben) dar und ist demnach zeit- und situationsbezogen kodiert.<sup>6</sup> Semantisches Wissen bezeichnet hingegen das Wissen über die Bedeutung einzelner Worte und Begriffe (etwa das Wissen, dass analytische Prüfungshandlungen Ergebnisprüfungen darstellen), ist zeit- und situationsunabhängig gespeichert und kann demnach auch als allgemeines Wissen bezeichnet werden.<sup>7</sup> In der Literatur wird semantisches Wissen häufig auch als Lexikon-Wissen bezeichnet, episodisches Wissen hingegen als autobiographisches Wissen.<sup>8,9</sup>

Von dem deklarativen Wissen ist das prozedurale Wissen abzugrenzen. Dieses kann als Wissen darüber bezeichnet werden, wie einzelne Handlungen, etwa die Durchführung einer analytischen Prüfungshandlung oder einer Risikobeurteilung, ausgeführt werden.<sup>10</sup> Diese Wissensart wird in der Literatur auch als „*Wissen, wie ...*“ oder *Handlungswissen* bezeichnet.<sup>1</sup> Ein entscheidender Vorteil prozeduralen Wissens ist darin zu sehen, dass die Handlungen scheinbar automatisch und mit weniger kognitivem Aufwand durchge-

---

<sup>1</sup> Für einen umfassenden Überblick über die in der Literatur vorliegenden Klassifizierungen von Wissen siehe ROMHARDT (2001), S. 49-50.

<sup>2</sup> Vgl. etwa KOONCE (1993), S. 60; MARTEN/QUICK/RUHNKE (2006), S. 394-395.

<sup>3</sup> Vgl. ARBINGER (1997), S. 18, 22.

<sup>4</sup> Vgl. ANDERSON (2001), S. 238.

<sup>5</sup> Vgl. ARBINGER (1997), S. 18.

<sup>6</sup> Vgl. SCHREIBER (2000), S. 194 sowie BRANDER/KOMPA/PELTZER (1989), S. 46.

<sup>7</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 46; ARBINGER (1997), S. 19.

<sup>8</sup> Vgl. etwa ZIMBARDO (1995), S. 327-328 oder JOHNSON (1994), S. 36-37.

<sup>9</sup> Allerdings kann sich eine strikte Trennung zwischen semantischem und episodischem Wissen als schwierig erweisen, da sie sich durchaus auf ein und denselben Aspekt beziehen können. So kann beispielsweise die Aussage, dass analytische Prüfungshandlungen effektive Maßnahmen zur Aufdeckung von Fehlern im Jahresabschluss darstellen, sowohl auf dem Wege der Mitteilung oder durch Lesen in einem Lehrbuch erworben worden sein, als auch durch eigene Erfahrung, also der Durchführung einer analytischen Prüfungshandlung im Rahmen eines Prüfungsauftrags. Im ersten Fall würde man diese Wissensseinheit als semantisches Wissen, im zweiten Fall als episodisches Wissen bezeichnen.

<sup>10</sup> Vgl. ANDERSON (2001), S. 238, 241-243; KOONCE (1993), S. 60.

führt werden.<sup>2</sup> Allerdings bedarf der Erwerb von prozeduralem Wissen der wiederholten Ausführung bzw. Übung und kann nicht durch reine Mitteilung angeeignet werden.<sup>3</sup> Zudem kann prozedurales Wissen nur schwer expliziert, also einem Dritten gegenüber ausgedrückt werden.<sup>4</sup>

Deklaratives und prozedurales Wissen müssen aber nicht unbedingt als ein sich ausschließendes Begriffspaar aufgefasst werden. Nach dem ACT\*-Modell von ANDERSON können die beiden Wissensarten auch die Endpunkte auf einer Wissensskala einnehmen.<sup>5</sup> Nach diesem Modell steht deklaratives Wissen zu Beginn eines jeden Lernprozesses. Durch häufiges Ausführen einer bestimmten Handlung wird durch die Phase der sog. *Wissenskompilation* prozedurales Wissen erworben.<sup>6</sup>

### **3.2.2.3.4.2 Wissensstrukturen**

#### **3.2.2.3.4.2.1 Eigenschaften von Wissensstrukturen**

Wissensstrukturen können als mentale Komplexe bezeichnet werden, in denen inhaltlich bzw. sachlich zusammengehörende Wissenseinheiten organisiert werden. In der Literatur wird deshalb auch häufig synonym der Begriff der *Wissensorganisation* verwendet.<sup>7, 8</sup>

Die wichtigste Eigenschaft von Wissensstrukturen ist in ihrer Vernetztheit zu sehen. Einzelne Wissenseinheiten sind zunächst ohne jegliche Verbindungen zueinander gespeichert. Reines Fach- bzw. Faktenwissen etwa, kann als derartiges (noch) unstrukturi-

---

<sup>1</sup> Vgl. ARBINGER (1997), S. 20.

<sup>2</sup> Vgl. SCHREIBER (2000), S. 241.

<sup>3</sup> Vgl. ARBINGER (1997), S. 22.

<sup>4</sup> Vgl. ANDERSON (2001), S. 238.

<sup>5</sup> „ACT“ steht für „Adaptive Control of Thought“, und das Sternchen bedeutet, dass es sich um eine erweiterte Version des ursprünglichen ACT-Modells handelt.

<sup>6</sup> Vgl. ANDERSON (1987), S. 192-210.

<sup>7</sup> Vgl. etwa FREDERICK/HEIMAN-HOFFMAN/LIBBY (1994), S. 1-2 oder TUBBS (1992), S. 783-784.

<sup>8</sup> In der Literatur ist auch häufig der Begriff der Wissensrepräsentation anzutreffen. Eine strikte Trennung zwischen Wissensstrukturen und Wissensrepräsentation ist dabei jedoch mit erheblichen Schwierigkeiten verbunden. Häufig werden unter den beiden Begriffen die gleichen Sachverhalte subsumiert. So werden semantische Netze, Produktionen, Prototypen, Schemata oder Skripte von einigen Autoren unter dem Oberbegriff Wissensrepräsentationsarten zusammengefasst (vgl. etwa REIMANN (1998), S. 341-350; OPWIS (1992), S. 49-101). Andere Autoren hingegen bezeichnen etwa Prototypen, Skripte und Schemata als unterschiedliche Arten von Wissensstrukturen (vgl. etwa SCHREIBER (2000), S. 198-231). Für die vorliegende Arbeit wird die Auffassung der zweiten Autorengruppe zugrunde gelegt. Wissensstrukturen und die Repräsentation von Wissen sind jedoch insofern miteinander verbunden, als dass im Rahmen der internen Repräsentation auf Wissensstrukturen durch Verknüpfung mit den Informationen des betrachteten (Problem-)Sachverhalts zurückgegriffen wird. Vgl. KOONCE (1993), S. 60; BEDARD/CHI (1993), S. 26; BIERSTAKER/BEDARD/BIGGS (1999), S. 19-20.

riertes Wissen bezeichnet werden.<sup>1</sup> FREDERICK bezeichnet die Anhäufung von reinem Faktenwissen als eine Art taxonomische Wissensorganisation, bei der einzelne Wissenseinheiten weder einer, noch unterschiedlicher hierarchischer Ebenen miteinander verbunden sind.<sup>2</sup> Ohne Vernetzung dieser einzelnen Einheiten bleibt dieses Wissen „träge“ und kann im Rahmen der Problemrepräsentation nur schwierig abgerufen werden, was den Erfolg der gesamten Problemlösung erheblich beeinträchtigt.<sup>3</sup> Erst durch aktive und problemorientierte Wissensanwendung, was dem Erwerb von episodischem bzw. Erfahrungswissen gleichkommt<sup>4</sup>, können die einzelnen Wissenseinheiten miteinander vernetzt und somit Wissensstrukturen erworben werden, auf die im Rahmen einer Problemrepräsentation zurückgegriffen werden kann.<sup>5</sup>

### **3.2.2.3.4.2.2 Typologien von Wissensstrukturen**

#### **3.2.2.3.4.2.2.1 Überblick**

Ähnlich wie bei den unterschiedlichen Wissensarten kann der einschlägigen Literatur auch eine Vielzahl an Vorschlägen zur Klassifizierung von Wissensstrukturen entnommen werden. So kann beispielsweise zwischen

- symbolischen und analogen Modellen,<sup>6</sup>
- wahrnehmungs- und bedeutungsbezogenen Modellen,<sup>7</sup>
- propositionalen, regelbasierten und analogen Modellen<sup>8</sup> sowie
- regel-, schema- und fallbasierten Modellen<sup>9</sup>

unterschieden werden, wobei die einzelnen Klassifikationen durch Überschneidungen gekennzeichnet sind. Welcher Typologisierung dabei zu folgen ist, hängt in erster Linie von den zu betrachteten Wissensarten ab. Daher sollen an dieser Stelle jene Modelle vorgestellt werden, die in besonderer Weise dazu geeignet sind, deklaratives und proze-

---

<sup>1</sup> Vgl. Hierzu auch Kap.3.2.2.3.4.1.

<sup>2</sup> Vgl. FREDERICK (1991), S. 241-242.

<sup>3</sup> Vgl. GRUBER/MANDL (1996), S. 18.

<sup>4</sup> Vgl. hierzu Kap. 3.2.2.3.4.1.

<sup>5</sup> Ähnlich argumentieren SHANTEAU (1993), S. 54; ROTHE/SCHINDLER (1996), S. 39; SCHREIBER (2000), S. 199-200.

<sup>6</sup> Vgl. etwa OPWIS (1992), S. 49-50 oder OPWIS/LÜER (1996), S. 347-348.

<sup>7</sup> Vgl. etwa ANDERSON (2001), S. 107-138 bzw. 139-172.

<sup>8</sup> Vgl. etwa ARBINGER (1997), S. 18-26.

<sup>9</sup> Vgl. REIMANN (1998), S. 343-353.

durales Wissen zu vernetzen und zu rekonstruieren. Dabei handelt es sich um propositionale, regelbasierte und schemabasierte Repräsentationsmodelle.<sup>1</sup> Während sich propositionale Modelle insbesondere dazu eignen, deklaratives Wissen zu modellieren, werden regelbasierte Modelle in der Regel herangezogen, um prozedurales Wissen zu rekonstruieren. Streng genommen lassen sich die einzelnen Wissensarten jedoch nicht eindeutig einem Modell zuordnen, da sie grundsätzlich unabhängig voneinander zu betrachten sind.<sup>2</sup> Schemabasierte Modelle können als eine Sonderform betrachtet werden, da sie sowohl Aspekte propositionaler als auch regelbasierter Modelle beinhalten.

### **3.2.2.3.4.2.2.2 Propositionale Modelle**

Propositionale Modelle finden insbesondere bei der Rekonstruktion deklarativen Wissens Verwendung. Eine Proposition stellt dabei die kleinste Wissenseinheit dar, die als eine selbstständige Aussage, d.h. von anderen Aussagen losgelöst, betrachtet werden kann.<sup>3</sup> Der Begriff der Proposition kommt dabei aus den Bereichen Logik und Linguistik und beschreibt den deklarativen Gehalt einer Aussage. Unter deklarativem Gehalt versteht man dabei den Bedeutungsgehalt einer Äußerung, der unabhängig von der exakten äußeren Darbietung eine eigenständige Behauptung darstellt, die sinnvoll als wahr oder falsch bezeichnet werden kann.<sup>4</sup> Sprachlich unterschiedliche Formulierungen, die jedoch den gleichen Bedeutungsgehalt aufweisen, lassen sich durch die gleiche Proposition darstellen. So können etwa die Sätze „eine analytische Prüfungshandlung stellt eine indirekte Prüfungsmethode dar“ bzw. „zu den indirekten Prüfungsmethoden zählen etwa analytische Prüfungshandlungen“ durch die Proposition

*Merkmal von (analytische Prüfungshandlung, indirekte Prüfungsmethode)*

wiedergegeben werden, wobei „analytische Prüfungshandlung“ und „indirekte Prüfungsmethode“ als Variablen bezeichnet werden, die durch die Proposition in Relation zueinander gesetzt werden.

---

<sup>1</sup> Für die Modellierung menschlicher Expertise haben sich dabei insbesondere regel- und schemabasierte Modelle als geeignet erwiesen. Vgl. etwa SCHREIBER (2000), S. 200-201 sowie 239-240; REIMANN (1998), S. 343-351.

<sup>2</sup> Vgl. ARBINGER (1997), S. 17.

<sup>3</sup> Vgl. ANDERSON (2001), S. 147.

<sup>4</sup> Vgl. OPWIS/LÜER (1996), S. 351.



Zusammenhängende Propositionen, also Aussagen, die sich auf die gleiche Variable beziehen, werden in der Literatur häufig auch *Konzepte* genannt.<sup>1</sup> So würde man etwa die beiden Aussagen „eine analytische Prüfungshandlung ist eine indirekte Prüfung“ und „eine analytische Prüfungshandlung ist eine Plausibilitätsbeurteilung“ dem Konzept *analytische Prüfungshandlung* zuordnen. Konzepte, die dem gleichen Realitätsbereich zugeordnet werden können, werden im Gedächtnis durch ein Netzwerk miteinander verbunden. In der Literatur wird hierfür häufig der Begriff des *semantischen Netzes* verwendet.<sup>2</sup> Semantische Netze stellen ein Modell der Organisation deklarativen Wissens dar, die eine zentrale Stellung für die mentale Repräsentation der Bedeutung von sprachlich-begrifflichen Strukturen einnehmen.<sup>3</sup> Allerdings darf der Begriff des semantischen Netzes nicht so verstanden werden, dass es sich hier ausschließlich um die Vernetzung semantischer Wissenseinheiten handelt. Vielmehr kann auch episodisches Wissen Bestandteil eines semantischen Netzes sein.<sup>4, 5</sup>

Die zusammenhängenden Konzepte eines semantischen Netzes sind durch logische Über- und Unterordnungen hierarchisch miteinander verbunden.<sup>6</sup> Die Gesamtheit der hierarchischen Ordnungsbeziehungen wird als *Taxonomie* bezeichnet. So ist beispielsweise das weiter oben angesprochene Konzept „analytische Prüfungshandlung“ einerseits dem Konzept „ergebnisorientierte Prüfungsmethoden“ untergeordnet und andererseits dem Konzept „Vorjahresvergleich“ übergeordnet (siehe Abbildung 6). Die Konzepte auf den verschiedenen hierarchischen Ebenen werden durch bestimmte Aussagen definiert.<sup>7</sup> So ist, wie bereits oben aufgeführt, dem Konzept „analytische Prüfungshandlung“ die Aussage „indirekte Prüfung“ und „Plausibilitätsbeurteilung“ zugeordnet, dem Konzept „Ergebnisprüfung“ die Aussage „Nachweis der Vollständigkeit und Richtigkeit“ und dem Konzept „Vorjahresvergleich“ die Aussagen „Jahresabschlussposition“ und „Vergleich Buchwert  $t_0$  und  $t_1$ “.

---

<sup>1</sup> Vgl. etwa BRANDER/KOMPA/PELTZER (1989), S. 46-47 oder ANDERSON (2001), S. 152.

<sup>2</sup> Vgl. etwa STRUBE/SCHLIEDER (1998), S. 504; ARBINGER (1997), S. 18 oder OPWIS (1992), S. 55.

<sup>3</sup> Vgl. OPWIS/LÜER (1996), S. 354.

<sup>4</sup> In diesem Zusammenhang sei auf die schwierige Unterteilung zwischen semantischem und episodischem Wissen verwiesen (siehe Kap. 3.2.2.3.4.1).

<sup>5</sup> In der Literatur wird z.T. sogar die Meinung vertreten, dass die Verbindungen zwischen einzelnen Wissenseinheiten, also die Wissensvernetzung, erst durch Erfahrungen vollzogen werden kann. Rein semantisches Wissen wird als „träge“ bezeichnet, da es in konkreten Fällen, etwa bei der Lösung von Problemen, nur bedingt zur Verfügung steht bzw. abgerufen werden kann. Ein derartig unvernetztes Wissen wird auch als „Wissenskompartimentalisierung“ bezeichnet. Vgl. SCHREIBER (2000), S. 199 m.w.N.

<sup>6</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 47.

<sup>7</sup> Vgl. ANDERSON (2001), S. 154.

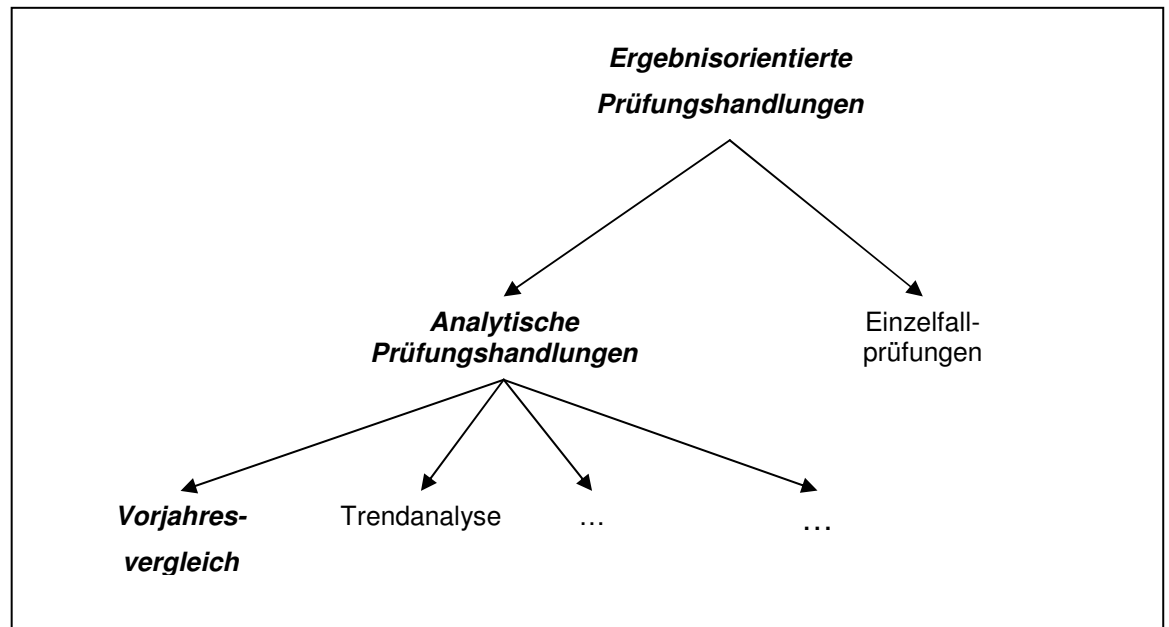


Abbildung 6: Taxonomie am Beispiel von ergebnisorientierten Prüfungshandlungen

Eine wichtige Annahme dieser propositionalen Modelle liegt darin, dass Aussagen, die für Konzepte auf höher liegenden Ebenen zutreffend sind, auch für darunter liegende Ebenen gelten. Diese wird in der Literatur auch als *kognitive Ökonomie*<sup>1</sup> bezeichnet, da hierdurch der Speicher- und Verarbeitungsaufwand reduziert wird. Demnach müsste man für die mentale Repräsentation bzw. Nachvollziehbarkeit des Satzes „ein Vorjahresvergleich ist eine indirekte Prüfung“ eine Schlussfolgerung (Inferenz) von einer Hierarchieebene in die unmittelbar darüberliegende ziehen. Soll der Satz „ein Vorjahresvergleich dient dem Nachweis der Vollständigkeit und Richtigkeit“ repräsentiert werden, sind bereits zwei Schlussfolgerungen notwendig.<sup>2</sup>

Zusammenfassend lässt sich festhalten, dass semantische Netze zwar dazu geeignet sind deklaratives Wissen zu repräsentieren, indem variablenbezogene Propositionen in Konzepten zusammengefasst werden und diese sowie deren Aussagen durch das hierarchische Vererbungsprinzip miteinander verknüpft werden. Allerdings reicht deklaratives Wissen bei weitem nicht dazu aus, um das prüferische Vorgehen bei bestimmten Prüfungsaufgaben, etwa analytischen Prüfungshandlungen, zu modellieren. Hierzu scheint es insbesondere notwendig, das Wissen über bestimmte Fertigkeiten, etwa Problemlö-

<sup>1</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 47.

<sup>2</sup> Den empirischen Nachweis derartiger Wissensrepräsentationen erbrachten COLLINS/QUILLIAN (1969), indem sie Probanden die Richtigkeit bzw. Falschheit von Aussagen beurteilen ließen. Dabei konnte festgestellt werden, dass desto mehr Zeit benötigt wurde, je mehr hierarchische Ebenen durchsucht bzw. Schlussfolgerungen gezogen werden mussten.

sungsfähigkeiten, zu berücksichtigen, die sich, wie bereits zuvor erwähnt, eher durch prozedurales Wissen rekonstruieren lassen.<sup>1</sup> Bei der Lösung von Problemen haben propositionale Modelle ihre Erklärungskraft eher in der ersten Repräsentation des Problem-sachverhalts, also im Verstehen des Problems.

### 3.2.2.3.4.2.2.3 Regelbasierte Modelle

Regelbasierte Modelle, mitunter auch als Produktionsregeln<sup>2</sup> oder Produktionssysteme<sup>3</sup> bezeichnet, sind insbesondere zur Rekonstruktion von prozeduralem Wissen geeignet. Produktionen stellen „die elaborierteste Rahmenkonzeption dar, in der Strukturtheorien des Gedächtnisses mit Annahmen der Informationsverarbeitung miteinander verbunden sind.“<sup>4</sup> Der Grund ihrer weitgehenden Akzeptanz ist u.a. darin zu sehen, dass sie menschliches Verhalten, insbesondere auch Problemlösungsverhalten, in geeigneter Weise rekonstruieren können.<sup>5</sup> Dazu trägt auch die einfache Architektur von regelbasierten Modellen bei, deren Bestandteile im Folgenden wiedergegeben werden sollen:<sup>6</sup>

- *Datenspeicher*

Der Datenspeicher enthält das dem System bekannte deklarative Wissen.

- *Produktionenspeicher*

Der Produktionenspeicher enthält das gesamte prozedurale Wissen in Form einer Menge von *Produktionen*. Einzelne Produktionen können als einfache Wenn-Dann-Beziehungen beschrieben werden, bei denen der Wenn-Teil genau die Bedingungen festlegt, die erfüllt sein müssen, um die Ausführung der in dem Dann-Teil gespeicherten Aktion auszulösen.<sup>7</sup> Dabei stehen die einzelnen Produktionen gleichwertig nebeneinander. D.h. sie sind unabhängig voneinander, so dass weder eine direkte In-

---

<sup>1</sup> Vgl. hierzu insbesondere Kap. 3.2.2.3.4.1.

<sup>2</sup> Vgl. etwa STRUBE/SCHLIEDER (1998), S. 506.

<sup>3</sup> Vgl. etwa OPWIS/LÜER (1996), S. 350, 374.

<sup>4</sup> OPWIS (1992), S. 73.

<sup>5</sup> NEWELL/SIMON (1972) führten Produktionen zu Beginn der siebziger Jahre in die Psychologie ein. In der Folgezeit war es insbesondere ANDERSON (1976, 1996, 2001), der die vielseitigen Möglichkeiten einer psychologischen Interpretation von Produktionen im Rahmen seines ACT\*-Modells aufzeigte und die gesamte kognitionspsychologische Theoriebildung wesentlich beeinflusste. Einen Überblick über psychologische Arbeiten, die sich auf regelbasierte Modellannahmen stützen, gibt OPWIS/LÜER (1996), S. 374.

<sup>6</sup> Die folgenden Ausführungen lehnen sich weitestgehend an OPWIS (1992), S. 73-74 an.

<sup>7</sup> Vgl. ARBINGER (1997), S. 21.

teraktion untereinander, noch der Aufruf einer Produktion durch eine andere Produktion möglich ist.<sup>1</sup>

- *Arbeitsspeicher*

Der Arbeitsspeicher enthält die interne Repräsentation der aktuellen (Problem-)Situation bzw. den aktiven Ausschnitt des semantischen Netzes. Hierzu bedient sich der Arbeitsspeicher des im Datenspeicher vorhandenen Wissens.

- *Interpreter*

Die Aufgabe des Interpreters liegt darin, die Informationsverarbeitung zu steuern. Dabei

- vergleicht er die im Arbeitsspeicher vorhandenen Informationen mit den Wenn-Teilen der vorhandenen Produktionen (*pattern matching*),<sup>2</sup>
- wählt er bei Mehrfachübereinstimmungen eine Produktion aus der Menge anwendbarer Produktionen aus (*conflict resolution*) und
- führt er die im Dann-Teil der ausgewählten Produktion gespeicherte Aktion aus (*execution*).

Im Rahmen der Modellierung des menschlichen Gedächtnisses bzw. der mentalen Repräsentation von prozeduralem Wissen werden dem Kurzzeitgedächtnis die Funktionen des Arbeitsspeichers sowie des Interpreters und dem Langzeitgedächtnis die Funktionen des Daten- und Produktionenspeichers zugeordnet.<sup>3</sup>

Neben diesen strukturtheoretischen Ähnlichkeiten weisen regelbasierte Modelle jedoch weitere Gründe auf, weshalb sie eine weitgehende Akzeptanz zur Modellierung der mentalen Repräsentation von Wissen erlangen konnten. Ein wichtiger Aspekt ist dabei, dass man mit ihnen das Vorgehen von Menschen bei der Lösung von Problemen in relativ einfacher Form modellieren kann.<sup>1</sup> Dabei wird angenommen, dass der Wenn-Teil die spezifischen Merkmale des (Problem-)Sachverhalts umfasst und der Dann-Teil die zur Bearbeitung des Problems notwendige Handlung.

Weitere Gründe für die Beliebtheit des Modells, sind darin zu sehen, dass man mit ihnen diverse psychologische Befunde, insbesondere im Bereich der Problemlösungs- und

---

<sup>1</sup> Vgl. OPWIS (1992), S. 81.

<sup>2</sup> Dass sich Produktionen dazu eignen, das Problemlösen als eine Mustererkennung zu beschreiben, zeigt auch REIMANN (1998), S. 344.

<sup>3</sup> Eine ähnliche Beschreibung findet sich bei OPWIS (1992), S. 83.

Expertiseforschung, erklären kann. So können Produktionen grundlegende kognitionspsychologische Erkenntnisse über parallel und seriell ablaufende Informationsverarbeitungsprozesse integrieren. In der Vergleichs- bzw. *pattern matching*-Phase läuft die Mustererkennung parallel ab, d.h. der Interpretier ist in der Lage, den Wenn-Teil mehrerer Produktionen gleichzeitig mit den Informationen des Arbeitsspeichers zu vergleichen. Hingegen sind Auswahl- bzw. *conflict resolution*-Phase und Ausführungs- bzw. *execution*-Phase durch eine serielle Informationsverarbeitung gekennzeichnet, d.h. Entscheidungen und Handlungen können nur nacheinander getroffen bzw. ausgeführt werden.<sup>2</sup>

Weiterhin eignen sich Produktionen dazu, sowohl vorwärts- als auch rückwärtsgerichtete Problemlösungsstrategien zu rekonstruieren.<sup>3</sup> Rückwärtsgerichtete bzw. zielgesteuerte Strategien basieren auf Produktionen, deren Wenn-Teil spezifische Zielinformationen umfasst, während vorwärtsgerichtete bzw. datengesteuerte Strategien durch Wenn-Teile ausgestattet sind, die keine Zielinformationen beinhalten.<sup>1</sup>

Zusammenfassend bleibt festzuhalten, dass regelbasierte Modelle gegenüber propositionalen Modellen den Vorteil aufweisen, dass sich mit ihnen nicht nur die mentale Erfassung des Problemsachverhalts erklären lässt, sondern auch die Suche im Problemraum sowie die Ausführungen von Operationen, die letztendlich die Transformation des Ausgangs- in den Zielzustand sicherstellen sollen. Weiterhin ermöglichen sie es, fundierte kognitionspsychologische Erkenntnisse, wie etwa Annahmen über seriell bzw. parallel ablaufende Informationsverarbeitungsprozesse oder Annahmen über vorwärts- bzw. rückwärtsablaufende Problemlösungsstrategien, in einer geeigneten Art und Weise nachvollziehen zu können. Daher scheinen sie besser als propositionale Modelle dazu in der Lage zu sein, die Grundlage für die Analyse der prüferischen Vorgehensweise bzw. Urteilsbildung zu bilden.

#### **3.2.2.3.4.2.2.4 Schemabasierte Modelle**

Wie bereits weiter oben dargestellt, handelt es sich bei schemabasierten Modellen – im Folgenden auch als *Schemata* bezeichnet – um eine Sonderform der Strukturierung von Wissen. Ihre Besonderheit liegt darin, dass sie einerseits Aspekte propositionaler Modelle erfüllen, indem sie ebenfalls auf die hierarchische Anordnung von Wissenseinhei-

---

<sup>1</sup> Vgl. REIMANN (1998), S. 343.

<sup>2</sup> Vgl. OPWIS/LÜER (1996), S. 384.

<sup>3</sup> Vgl. hierzu insbesondere Kap 3.2.2.3.3.2.

ten abstellt und somit Eigenschaften eines Netzes erfüllt. Andererseits geht sie jedoch weit über die reine Repräsentation von deklarativem Wissen hinaus, da sie auch zur Rekonstruktion von prozeduralem Wissen herangezogen werden kann. Darüber hinaus lassen sich mit Schemata diverse Aspekte des Problemlösungsverhaltens von Individuen erklären, was durch einfache regelbasierte Modelle nicht oder nur bedingt möglich ist.

Schemata können als Wissenskomplexe über typische Zusammenhänge in einem Realitätsbereich bezeichnet werden.<sup>2</sup> Hervorzuheben ist allerdings, dass es sich bei Schemata nicht um logische Zusammenhänge oder Definitionen, sondern um verallgemeinerte Erfahrungen handelt.<sup>3</sup> Damit weisen Schemata zunächst einen starken Bezug zu episodischem bzw. Erfahrungswissen auf. Der wesentliche Unterschied ist allerdings in dem Allgemeinsgrad zu sehen. Während episodisches Wissen per definitionem kontext- bzw. situationsabhängig gespeichert wird, stellen Schemata verallgemeinerte Erfahrungen dar; sie sind also abstrakter.<sup>4</sup>

Episodisches Wissen bildet allerdings die Grundlage für die Entwicklung von Schemata. Sie stehen zu Beginn eines jeden Schemaerwerbs. Ihre Dekontextualisierung erfolgt durch wiederholte Erfahrungen mit ähnlichen Situationen, in deren Verlauf die neuen situativen Informationen in den zunächst rudimentären Schemata integriert werden. Durch wiederholte Generalisierungs- und Abstraktionsprozesse verlieren die spezifischen Erfahrungen ihre kontextuellen Merkmale und erhalten einen hohen Allgemeinsgrad.<sup>5</sup> Schemata repräsentieren demnach ein allgemeines Modell oder eine Grundstruktur für einen bestimmten Realitätsbereich und nicht einen spezifischen Sachverhalt.<sup>6</sup> Allerdings lassen sich durch Schemata Schlussfolgerungen über spezifische Sachverhalte treffen.<sup>7</sup>

Ein Schema besteht aus zwei Komponenten. Zum einen umfasst es Merkmale einer (allgemeinen) Problemsituation und zum anderen Informationen über das anzuwendende Problemlösungsverhalten.<sup>8</sup> Damit weisen einzelne Schemata zumindest prinzipiell betrachtet eine starke Ähnlichkeit zu Produktionen auf. Allerdings sind Produktionen darauf ausgelegt, nach direkten Übereinstimmungen zwischen Merkmalen des Problem-

---

<sup>1</sup> Vgl. OPWIS (1992), S. 82.

<sup>2</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 44.

<sup>3</sup> Vgl. WIPPICH (1984), S. 57-58.

<sup>4</sup> Vgl. WIPPICH (1984), S. 80.

<sup>5</sup> Vgl. GRUBER/MANDL (1996), S. 25.

<sup>6</sup> Vgl. ANDERSON (2001), S. 157 sowie WIPPICH (1984), S. 57.

<sup>7</sup> Vgl. ANDERSON (2001), S. 157.

<sup>8</sup> Vgl. REIMANN (1998), S. 347.

sachverhalts und den Informationen des Wenn-Teils zu suchen. Fällt dieser Abgleich jedoch komplizierter aus, etwa weil lediglich eine mittelbare Übereinstimmung vorliegt und ein erfolgreiches *pattern matching*<sup>1</sup> das Ziehen von Schlussfolgerungen bedarf, so stoßen Produktionen schnell an ihre Grenzen. Schemata bringen demgegenüber den Vorteil mit sich, dass sie aufgrund ihrer abstrakten Natur auch mit mittelbaren Übereinstimmungen auskommen.<sup>2</sup>

Weiterhin zeichnen sich Schemata gegenüber Produktionen dadurch aus, dass sie in vielfältigen Relationen zueinander stehen können. Ähnlich wie bei semantischen Netzen sind auch Schemata hierarchisch angeordnet, wobei aus übergeordneten Schemata die Informationen für jeweils untergeordnete Schemata abgeleitet werden können.<sup>3</sup> Demnach können Schemata in einer Spezialisierungs-Generalisierungs-Relation zueinander stehen. Dieses hat wiederum weitgehende Implikationen für die Phase des Musterabgleichs, da Schemata auf diesem Wege auch mit einer partiellen Übereinstimmung auskommen. Schemata mit einem hohen Generalisierungsgrad decken dabei immer nur teilweise die Aspekte eines zu bearbeitenden Problems ab, steuern das Problemverständnis jedoch in die richtige Richtung, nämlich zu den zugehörigen spezielleren Schemata, die probeweise aktiviert werden können, um ein Problem zu verstehen und lösen zu können.<sup>4</sup> Demnach können hierarchisch angeordnete Schemata als eine Art Diskriminanzbaum bezeichnet werden, der es dem Problemlöser ermöglicht, Klassifikationsleistungen zu erbringen.<sup>5, 6</sup>

Eine weitere Eigenschaft von Schemata liegt darin, dass die in ihr gespeicherten Merkmale einer Problemsituation sowohl variable als auch fixe *slots* umfassen können. Variable slots werden mit den konkreten Informationen des betrachteten Problemsachverhalts gefüllt, fixe slots sind typische Wertebelegungen (sog. *Default*-Werte), also Annahmen über Werte in typischen Situationen.<sup>7</sup> Wenn demnach eine spezifische Problemsituation ein situationsangemessenes Schema aktiviert, so wird das abstrakte Sche-

---

<sup>1</sup> Vgl. hierzu Kap. 3.2.2.3.4.2.2.3.

<sup>2</sup> Vgl. REIMANN (1998), S. 348.

<sup>3</sup> Vgl. ANDERSON (2001), S. 157.

<sup>4</sup> Vgl. REIMANN (1998), S. 349.

<sup>5</sup> Vgl. REIMANN (1998), S. 348.

<sup>6</sup> In der Literatur wird die Meinung vertreten, dass Schemata nicht zwingend in einer Generalisierungs-Spezialisierungs-Relation angeordnet sein müssen. Vielmehr kann die hierarchische Anordnung auch nach einer Teil-Ganzes-Relation erfolgen. So etwa ANDERSON (2001), S. 157 oder REIMANN (1998), S. 347. Derartige Schemahierarchien sind zu den gleichen Klassifikationsleistungen fähig, da auch sie mit partiellen Musterübereinstimmungen auskommen und die weitere Informationsverarbeitung in die darunter liegende Ebene leiten.

<sup>7</sup> Vgl. ANDERSON (2001), S. 156-157 sowie REIMANN (1998), S. 349.

ma durch konkrete situative Merkmale präzisiert, was den Problemlöser in die Lage versetzt, ein kontextspezifisches Problemverständnis aufzubringen. Fehlende Informationen werden durch Default-Werte ersetzt, die schemakonsistentes allgemeines Wissen repräsentieren.<sup>1</sup> Default-Werte können demnach als plausibles Schließen bezeichnet werden, was überall dort von großer Bedeutung ist, wo kein vollständiges Wissen vorhanden ist.<sup>2</sup> Hierzu zählt sicherlich auch der gesamte Bereich der Jahresabschlussprüfung.

Zusammenfassend kann also gesagt werden, dass ein einmal aktiviertes angemessenes Schema dem Problemlöser einen Deutungsrahmen zur Verfügung stellt, der es ihm erlaubt, ein adäquates Problemverständnis aufzubauen.<sup>3</sup> Durch das einmal aktivierte Schema wird die weitere Problembearbeitung wesentlich beeinflusst. So leitet etwa das aktivierte Schema die Aufmerksamkeit des Problemlösers auf die problem- und lösungsrelevanten Informationen, die durch das jeweilige Schema nahe gelegt werden, steuert die Interpretation des Problemsachverhalts und erlaubt es dort Schlussfolgerung zu ziehen, wo Informationen zum Problemsachverhalt unvollständig oder mehrdeutig sind.<sup>4</sup> Insgesamt stellen Schemata somit das differenzierteste Modell zur Strukturierung von Wissen dar, mit denen diverse Aspekte des Problemlösungsverhaltens von Individuen angemessen berücksichtigt werden können. Insbesondere die Annahmen über den hohen Allgemeinheitsgrad, der hierarchischen Anordnung und den Default-Werten erlauben es, Informationsverarbeitungsprozesse zu modellieren, die mit den flachen Strukturen von Produktionen und dessen hohen Konkretisierungsgrad nicht erklärt werden können.

#### **3.2.2.4 Auswahl eines verhaltensorientierten Modells zur Abbildung der prüferischen Vorgehensweise**

Die vorherigen Ausführungen zeigen, dass das Informationsverarbeitungs- bzw. Problemlösungsmodell wohl das am weitesten elaborierte unter den vorgestellten Modellen darstellt. Diesem ist wohl das größte Potenzial zur realitätsnahen Darstellung und Analyse der prüferischen Vorgehensweise bei der Bearbeitung eines prüfungsbezogenen

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 206.

<sup>2</sup> Vgl. REIMANN (1998), S. 349.

<sup>3</sup> Vgl. BRANDER/KOMPA/PELTZER (1989), S. 44-45.

<sup>4</sup> Vgl. CHOO (1989), S. 481.



Problemsachverhalts zu attestieren.<sup>1</sup> Sein Anspruch, reale Urteilsbildungsprozesse erklären zu wollen, macht es prädestiniert dafür, die prüferische Urteilsbildung aus einer verhaltensorientierten Sichtweise zu analysieren. Ein wesentlicher Vorteil dieses Modells ist vor allem darin zu sehen, dass eine Analyse des Urteils nicht wie bei der starren Input-Output-Betrachtung der Attributions- und Regressionsmodelle vor dem Hintergrund deduktiver Modellfestlegungen, sondern durch induktive Ableitung des Modells erfolgt.<sup>2</sup> Ein auf dem Informationsverarbeitungsansatz basierendes Problemlösungsmodell kann somit mehr „Licht in den Black Box ‚Prüfer‘ bringen“<sup>3</sup>, indem es die bei der prüferischen Urteilsbildung stattfindenden Prozesse konkret verfolgt. Um jedoch ein größtmögliches Verständnis über die prüferische Urteilsbildung zu erlangen, soll im weiteren Verlauf dieser Arbeit nicht ausschließlich auf den Informationsverarbeitungsansatz abgestellt werden. Vielmehr wird auch eine Integration der aus den Modellen subjektiver Wahrscheinlichkeitsschlüsse hervorgegangen Erkenntnisse der „heuristics-and-biases“-Forschung angestrebt.<sup>4</sup>

Auf dieser Grundlage wird im Folgenden ein Ablaufmodell konzipiert, welches die prüferische Urteilsbildung bei der Durchführung von analytischen Prüfungshandlungen möglichst exakt rekonstruieren soll. Dabei wird auf das unter Kap. 2.3.2 skizzierte Ablaufmodell analytischer Prüfungshandlungen zurückgegriffen. Aus dem Blickwinkel eines realwissenschaftlichen Theorieverständnisses soll das Ablaufmodell zudem den Ansprüchen des in Kap. 3.2.1.2.2 deduzierten Forschungsprogramms genügen.

---

<sup>1</sup> So auch GANS (1986), S. 197.

<sup>2</sup> Vgl. HALLER (1992), S. 417.

<sup>3</sup> SCHREIBER (2000), S. 14.

<sup>4</sup> Ähnlich argumentiert SCHREIBER (2000), S. 14.

### 3.3 Entwicklung eines Ablaufmodells zur Durchführung von analytischen Prüfungshandlungen aus dem Blickwinkel eines realwissenschaftlichen Theorieverständnisses

Wie dem in Kap. 3.2.1.2.2 abgeleiteten Forschungsprogramm zu entnehmen ist, ist es in einem *ersten Schritt* zunächst notwendig, den prüfungsbezogenen Sachverhalt genauer zu spezifizieren. Bei der vorliegenden Untersuchung handelt es sich um die Durchführung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung. Was unter analytischen Prüfungshandlungen zu verstehen ist und wie die prüferische Vorgehensweise im Rahmen von analytischen Prüfungshandlungen aussieht, wurde in den Kap. 2.1.1 und 2.3.2 ausführlich erläutert. Demnach beginnt die Durchführung einer analytischen Prüfungshandlung mit der *Bildung eines Erwartungswerts*. Hierzu bedient sich der Prüfer den unterschiedlichen zur Verfügung stehenden Methoden analytischer Prüfungshandlungen.<sup>1</sup> Anschließend *legt* er die *Toleranzgrenzen fest*, um daraufhin den tatsächlich ausgewiesenen Buchwert und den entwickelten Erwartungswert im Sinne eines *Soll-Ist-Vergleichs* gegenüberzustellen. Befindet sich der Buchwert innerhalb der festgelegten Toleranzgrenzen, so kann der Prüfer dieses zum Anlass nehmen, das zugrunde liegende Prüffeld als normenkonform anzunehmen oder den geplanten Prüfungsumfang zu reduzieren. Liegt der Buchwert hingegen außerhalb der Toleranzgrenzen, obliegt es dem Prüfer, diese Abweichung zu *beurteilen*. Hierzu wird er zunächst eine *Hypothese* über die Abweichungsursache *generieren*. Anschließend wird er weitere auf die generierte Hypothese bezogene *Informationen beschaffen*, um daraufhin seine *Hypothese* vor dem Hintergrund der erlangten Erkenntnisse zu *bewerten*. Die weitere Vorgehensweise des Prüfers ist vom Ergebnis der Beurteilung abhängig. Konnte die Hypothese etwa nicht bestätigt werden, wird er diese verwerfen und eine neue Hypothese generieren. Erweist sich die ursprünglich generierte Hypothese hingegen als korrekt, wird der Prüfer schließlich eine *Prüfungsentscheidung* treffen. Der Prüfer könnte etwa, falls sich eine Fehlerhypothese bestätigt hat, eine Nachbuchung des Mandanten veranlassen. Hat sich hingegen eine Nicht-Fehlerhypothese verifiziert, könnte der Prüfer das Prüffeld etwa als normenkonform annehmen, weitere auf das gleiche Prüffeld gerichtete Prüfungshandlungen anordnen oder den Prüfungsumfang einschränken.

Nachdem man sich auf einen zu betrachtenden Sachverhalt, hier die Durchführung von analytischen Prüfungshandlungen, geeinigt hat, ist in einem *zweiten Schritt* zu klären,

---

<sup>1</sup> Vgl. hierzu ausführlich Kap. 2.2.

welches verhaltensorientierte Modell möglichst umfassend den Eigenheiten des zuvor beschriebenen Ablaufs Rechnung trägt. Um den hohen Ansprüchen einer sowohl als Real- als auch angewandten Wissenschaft verstandenen Prüfungslehre gerecht zu werden, sollte das heranzuziehende Modell eine geeignete Grundlage zur Erlangung eines umfassenden Verständnisses über die prüferische Vorgehensweise bilden. Hierzu sind insbesondere Modelle zu präferieren, die sich weitestgehend von einer einfachen „black box“-Betrachtung lösen. Die heranzuziehenden Modelle sollten den betrachteten Sachverhalt möglichst detailliert und unter Berücksichtigung der relevanten Einflussfaktoren rekonstruieren, um auf dieser Basis eine Vielzahl von theoriegeleiteten Aussagen über Ursache-Wirkungszusammenhänge eruieren zu können, welche die Grundlage für die Ableitung von auf die Gestaltung der Prüfungspraxis gerichteten Technologien bilden.

Wie unter Kap. 3.2.2.4 bereits festgestellt wurde, ist in dem Informationsverarbeitungsansatz nach NEWELL/SIMON ein Modell zu sehen, welches diesen umfassenden Anforderungen genügen kann. Neben der Entwicklung von prüfungsbezogenen Problemlösungsmodellen bietet dieses Modell insbesondere auch die Möglichkeit, Wissen und Wissensstrukturen als zentrale Einflussfaktoren in die Analyse zu integrieren. Daher soll im Folgenden ein Ablaufmodell gebildet werden, welches die Durchführung von analytischen Prüfungshandlungen als einen Problemlösungsprozess wiedergibt.

Folglich beginnt der Prüfer zunächst mit dem Aufbau einer internen Repräsentation des betrachteten Problems, welche im Folgenden als *Problemrepräsentation* bezeichnet wird. Im Rahmen von analytischen Prüfungshandlungen umfasst die Problemrepräsentation die Erwartungsbildung, die Festlegung von Toleranzgrenzen und die Durchführung des Soll-Ist-Vergleichs. Allerdings ist die Erwartungsbildung bereits durch das kognitiv weniger aufwendige Abrufen des deklarativen sowie prozeduralen Wissens möglich.<sup>1</sup> Als deklaratives Wissen<sup>2</sup> ließe sich in diesem Fall die Kenntnis des Prüfers darüber bezeichnen, was unter einer analytischen Prüfungshandlung zu verstehen ist und welche Methoden hierzu existieren. Darüber hinaus können auch rechnungsspezifische sowie mandanten- oder branchenbezogene Kenntnisse zum deklarativen Wissen des Prüfers zugeordnet werden.<sup>3</sup> Prozedurales Wissen<sup>1</sup> hingegen versetzt den Prüfer in die Lage, eine analytische Prüfungshandlung zur Beurteilung der Normenkonformität der Jahresabschlussposition durchzuführen, da es das sog. Ablaufwis-

---

<sup>1</sup> Vgl. KOONCE (1993), S. 59-60.

<sup>2</sup> Vgl. hierzu ausführlich Kap 3.2.2.3.4.1.

<sup>3</sup> So auch BIERSTAKER/BEDARD/BIGGS (1999), S. 20.

sen zur Durchführung einer Handlung bereitstellt.<sup>2</sup> Setzt man einmal voraus, dass ein Großteil der in der Jahresabschlussprüfung tätigen Personen zumeist eine intensive Ausbildung, etwa durch den Abschluss eines universitären Studiengangs, in Anspruch genommen, und für gewöhnlich in den ersten Monaten ihres Berufseinstiegs an zahlreichen Schulungsmaßnahmen teilgenommen haben, kann davon ausgegangen werden, dass sich der Prüfer durch die Entwicklung eines Erwartungswerts kaum einem wirklichen Problem gegenüber gestellt sieht.<sup>3</sup>

Bei der Festlegung von Toleranzgrenzen hängt die Komplexität der aufzubauenden Problemrepräsentation von dem eingesetzten Verfahren ab. Sollte der Prüfer auf eine Regressionsanalyse zurückgreifen, so steht ihm mit der Standardabweichung bereits ein Maß zur Festlegung der Toleranzgrenzen zur Verfügung.<sup>4</sup> Ebenso unproblematisch gestaltet sich die Toleranzfestlegung, wenn der Prüfer die 5%- bzw. 10%-Regel verwenden sollte.<sup>5</sup> Probleme bei der Festlegung von Toleranzgrenzen ergeben sich hingegen, wenn der Prüfer auf eine sog. *judgmental method* zurückgreift.<sup>6</sup> Hierbei kommt es beim Aufbau einer adäquaten Problemrepräsentation entscheidend darauf an, dass der Prüfer über das relevante Wissen hinsichtlich der Kriterien zur Festlegung von Toleranzgrenzen verfügt und dieses auch abrufen kann.

Während des anschließenden Soll-Ist-Vergleichs hängen die an den Prüfer gestellten Ansprüche im Wesentlichen davon ab, ob eine wesentliche Abweichung zwischen dem gebildeten Erwartungswert und dem zu prüfenden Buchwert identifiziert wurde oder nicht. Sollte der Soll-Ist-Vergleich keine wesentliche Abweichung anzeigen, so ist die Problemrepräsentation abgeschlossen. Ungleich komplexer gestaltet sich hingegen der Aufbau einer adäquaten Problemrepräsentation, falls das Ergebnis des Soll-Ist-Vergleichs eine unerwartete Abweichung signalisiert.<sup>7, 8</sup> In diesem Fall sollte die Problemrepräsentation den Prüfer möglichst in die Lage versetzen, ein differenziertes Verständnis über den zugrunde liegenden Sachverhalt zu erlangen, und ihm Möglichkeiten für die Lösung des Problems bieten. Bei analytischen Prüfungshandlungen sollte die Lösung des Problems darin bestehen, potenzielle Erklärungsursachen und notwendige

---

<sup>1</sup> Vgl. hierzu ausführlich Kap. 3.2.2.3.4.1.

<sup>2</sup> Ähnlich MARTEN/QUICK/RUHNKE (2006), S. 394-395.

<sup>3</sup> Allerdings zeigen die Ergebnisse der empirischen Prüfungsforschung, dass Prüfer auch hierbei Verzerrungen unterliegen können. Siehe hierzu Kap. 3.5.2.2.2.

<sup>4</sup> Vgl. hierzu Kap. 2.2.4.2.3.

<sup>5</sup> Vgl. hierzu Kap. 2.3.2.2.2.

<sup>6</sup> Vgl. hierzu Kap. 2.3.2.2.2.

<sup>7</sup> Vgl. KOONCE (1993), S. 60.

<sup>8</sup> Hier spricht man auch von *red flags*. Vgl. etwa FEROZ/KIM/RAAB (2005), S. 17.

Aktivitäten zu deren Verifizierung oder Falsifizierung bereitzustellen.<sup>1</sup> Demnach stellt die Problemrepräsentation die Grundlage für das gesamte weitere prüferische Vorgehen dar und nimmt eine entscheidende Bedeutung für den Erfolg der durchgeführten Prüfungshandlung ein.<sup>2</sup>

Da prüfungsbezogene Sachverhalte in der Regel äußerst komplex sind und der Prüfer einer *begrenzten Rationalität*<sup>3</sup> unterliegt, kann der Prüfer zu Beginn jedoch nicht alle problemrelevanten Informationen verarbeiten. Vielmehr konstruiert er zunächst ein begrenztes subjektives Abbild des Problemsachverhalts.<sup>4</sup> Dieses unvollständige Problemverständnis beeinflusst maßgeblich die weitere Vorgehensweise des Prüfers bzw. die Wahl der anzuwendenden Problemlösungsstrategie.<sup>5</sup> Um das eingeschränkte Abbild zu vervollständigen, wird sich der Prüfer zunächst um die Beschaffung weiterer Informationen bemühen. Aus zeitlichen und kognitiven Restriktionen wird der Prüfer jedoch nicht jegliche Informationen einholen und verarbeiten. Vielmehr wird er seine Informationssuche auf einen bestimmten Aspekt des Problemsachverhalts fokussieren.

Eine Problemlösungsstrategie, unter deren Zuhilfenahme diese prüferische Vorgehensweise adäquat berücksichtigt wird, ist in der rückwärtsgerichteten *generate and test*-Strategie zu sehen.<sup>6</sup> Diese auch als Novizenstrategie bezeichnete Vorgehensweise geht explizit von einer anfänglich noch rudimentären und vagen Problemrepräsentation aus, die erst im Laufe der Problembewältigung vervollständigt bzw. ergänzt wird.<sup>7</sup> Hierbei erfolgt zunächst eine auf eine mögliche Abweichungsursache gerichtete *Hypothesengenerierung*, auf deren Grundlage die anschließende *Informationssuche* gesteuert wird. In der Literatur bezeichnet man dieses Vorgehen häufig auch als *hypothesengesteuerten Suchprozess*.<sup>8</sup> Allerdings ist zu berücksichtigen, dass das prüferische Vorgehen nicht zwangsläufig rückwärtsgerichtet ausgestaltet sein muss. Vielmehr hängt die konkrete Vorgehensweise von dem jeweiligen Wissensstand des Prüfers ab. Verfügt er etwa über ein umfassendes problemrelevantes Wissen, könnte das prüferische Vorgehen auch vorwärtsgerichtet ausgestaltet sein.<sup>9</sup> Bei diesem auch als Expertenstrategie bezeichneten

---

<sup>1</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 19.

<sup>2</sup> So auch SCHREIBER (2000), S. 151; KOONCE (1993), S. 60.

<sup>3</sup> Siehe hierzu ausführlich Kap. 3.2.2.2.1.

<sup>4</sup> Vgl. MARTEN/QUICK/RUHNKE (2007), S. 50-51.

<sup>5</sup> Siehe hierzu ausführlich Kap. 3.2.2.3.

<sup>6</sup> Siehe hierzu ausführlich Kap. 3.2.2.3.2.2.

<sup>7</sup> Ähnlich argumentieren auch MARTEN/QUICK/RUHNKE (2006), S. 395.

<sup>8</sup> So auch GANS (1986); RUHNKE (2000), S. 291; MARTEN/QUICK/RUHNKE (2007), S. 50-51; BIGGS/MOCK/QUICK (2000), S. 171; KOONCE (1993), S. 65; SCHREIBER (2000), S. 82-83.

<sup>9</sup> Siehe hierzu ausführlich Kap. 3.2.2.3.2.3.

Vorgehen konzentriert sich der Prüfer von vornherein auf die wichtigen Aspekte und Informationen des Problemsachverhalts und generiert Hypothesen auf Basis der gegebenen Fakten.<sup>1</sup> Da im Rahmen eines Problemlösungsprozesses jedoch auch unterschiedliche Problemlösungsstrategien zur Anwendung kommen können, ist es durchaus denkbar, dass der Prüfer zunächst mit einer vorwärtsgerichteten Suchstrategie beginnt, seine Hypothese im Lichte zusätzlich erlangter Informationen verwirft und sich anschließend aufgrund des nun eingetretenen mangelnden Problemverständnisses einer rückwärtsgerichteten bzw. hypothesengesteuerten Suchstrategie bedienen muss.

Die Kombination von Problemlösungsstrategien ist jedoch nicht ausschließlich auf den Suchprozess beschränkt. Will man etwa die Hypothesengenerierung auf eine umfassendere theoretische Fundierung abstellen, scheint es durchaus sinnvoll, weitere Problemlösungsstrategien in das Modell zu integrieren. Im einfachsten Fall könnten Hypothesen etwa anhand der Versuch-Irrtum-Strategie<sup>2</sup> generiert werden. Weitere Möglichkeiten zur Entwicklung von Abweichungshypothesen stellen etwa die Analogiebildung<sup>3</sup>, die Abstraktion<sup>4</sup> sowie die Verfügbarkeits- und Repräsentativitätsheuristik<sup>5</sup> nach KAHNEMAN/TVERSKY dar.

Unabhängig davon, ob die bisherige Vorgehensweise vorwärts- oder rückwärtsgerichtet ausgestaltet wurde, wird der Prüfer in einem nächsten Schritt seine Hypothesen vor dem Hintergrund der erlangten Informationen bewerten. Da sich die *Hypothesenbewertung* mit der Wahrscheinlichkeitsbeurteilung der zuvor generierten Hypothese bzw. Hypothesen beschäftigt, erscheint es durchaus denkbar, dass der Prüfer zur Komplexitätsreduktion auch auf die Wahrscheinlichkeitsheuristiken nach KAHNEMAN/TVERSKY zurückgreift.<sup>6</sup>

Der Problemlösungsprozess gilt dann als abgeschlossen, „wenn anhand der vorliegenden Informationen der Überzeugungsgrad einen Schwellenwert (Abbruchkriterium) erreicht, der die Formulierung eines normenkonformen *Prüfungsurteils* zulässt.“<sup>1</sup>

Charakteristisch für das deduzierte Problemlösungsmodell ist, dass es sich nicht um einen linearen Prozess, sondern um einen rückgekoppelten, iterativen Prozess handelt, bei

---

<sup>1</sup> In der empirischen Prüfungsforschung beschäftigen sich nur wenige Studien explizit mit der vorwärtsgerichteten Strategie. Vgl. O'DONNELL/LEHMAN (1999) sowie O'DONNELL (2004).

<sup>2</sup> Vgl. hierzu Kap. 3.2.2.3.3.2.1.

<sup>3</sup> Vgl. hierzu Kap. 3.2.2.3.3.2.5.3.

<sup>4</sup> Vgl. hierzu Kap. 3.2.2.3.3.2.5.3.

<sup>5</sup> Vgl. hierzu Kap. 3.2.2.2.2.2.

<sup>6</sup> Ähnlich argumentieren MARTEN/QUICK/RUHNKE (2007), S. 50-51.

dem eine Prozessphase zudem auch übersprungen werden kann.<sup>2</sup> Die Rückkopplungen sind umso stärker ausgeprägt, je schlechter sich das anfängliche Problemverständnis des Prüfers darstellt.<sup>3</sup>

Zusammenfassend bleibt festzuhalten, dass der Informationsverarbeitungsansatz eine umfassende theoretische Fundierung des auf die Durchführung von analytischen Prüfungshandlungen bezogenen Problemlösungsmodells ermöglicht. Es wurde gezeigt, dass das allgemeine Ablaufmodell analytischer Prüfungshandlungen vollends in das Problemlösungsmodell integriert werden konnte. Darüber hinaus konnte aufgezeigt werden, dass mittels Kombination der aus dem Informationsverarbeitungsansatz abgeleiteten Problemlösungsstrategien zusätzliches Erklärungspotenzial für das prüferische Vorgehen bei analytischen Prüfungshandlungen gewonnen werden konnte. Die nachfolgende Abbildung gibt das Problemlösungsmodell graphisch wieder.

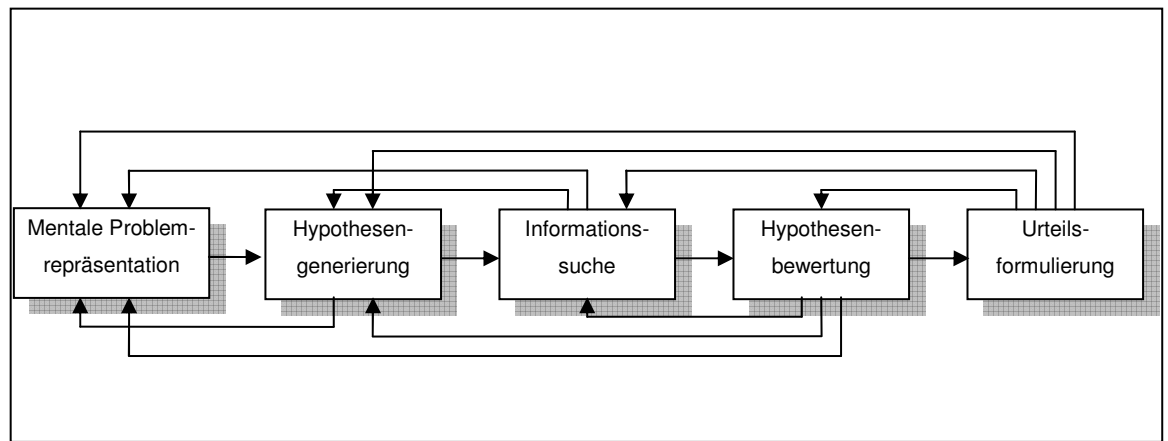


Abbildung 7: Ablaufmodell der Durchführung einer analytischen Prüfungshandlung aus dem Blickwinkel eines realwissenschaftlichen Theorieverständnisses<sup>4</sup>

<sup>1</sup> RUHNKE (2000), S. 291.

<sup>2</sup> Vgl. MARTEN/QUICK/RUHNKE (2006), S. 395; SCHREIBER (2000), S. 82; KOONCE (1993), S. 58-59.

<sup>3</sup> Vergleichbar argumentieren MARTEN/QUICK/RUHNKE (2006), S. 395.

<sup>4</sup> In Anlehnung an KOONCE (1993), S. 59.

### **3.4 Empirische Forschungsmethoden: Methoden der Datengewinnung und -auswertung**

#### **3.4.1 Vorbemerkungen**

Als empirische Forschungsmethoden bezeichnet man die systematische Vorgehensweise, um auf die Realität bezogene Erkenntnisse zu erlangen.<sup>1</sup> Dabei wird generell zwischen den Methoden der *Datengewinnung* und *Datenauswertung* differenziert.<sup>2</sup> Bezüglich des in Kap. 3.2.1.2.2 skizzierten erfahrungswissenschaftlichen Forschungsprogramms handelt es sich bei der Datengewinnung um den *vierten Schritt* und bei der Datenauswertung um den *fünften Schritt*.

Bei den Methoden der Datengewinnung kann zunächst eine Unterscheidung zwischen der *Befragung*, der *Beobachtung* und der *Inhaltsanalyse* vorgenommen werden.<sup>3</sup> Im Rahmen der Datenauswertung ist generell zwischen der *deskriptiven* und der *schließenden Statistik* zu differenzieren.<sup>4</sup>

#### **3.4.2 Methoden der Datengewinnung**

##### **3.4.2.1 Befragung**

Die Befragung gilt als eine der am häufigsten eingesetzten Methode in der empirischen Sozialforschung.<sup>5</sup> Dabei können sie generell in standardisierter, teil-standardisierter und nicht-standardisierter Form durchgeführt werden.<sup>6</sup> Bei *standardisierten* Befragungen liegen für alle Befragten die gleichen Fragen in der gleichen Reihenfolge vor.<sup>7</sup> Unter *teil-standardisierten* Befragungen versteht man solche Verfahren, die bestimmte Vorgaben umfassen, den Beteiligten jedoch bestimmte Freiräume in der konkreten Ausgestaltung der Befragung gewähren. Von einer *nicht-standardisierten* oder offenen Befra-

---

<sup>1</sup> Vgl. RUHNKE (2000), S. 333. Für eine Definition des Begriffs Methode sowie ihre Abgrenzung zur Methodologie wird insbesondere auf RUHNKE (2000), S. 233 Fn. 1 verwiesen.

<sup>2</sup> Vgl. etwa SCHNELL/HILL/ESSER (2005), S. 319-422, 441-474; ATTESLANDER (2006), S. 65-302 und SEIPEL/RIEKER (2003), S. 135-212.

<sup>3</sup> Vgl. SCHNELL/HILL/ESSER (2005), S. 319 sowie STIER (1996), S. 163.

<sup>4</sup> Vgl. RUHNKE (2000), S. 235.

<sup>5</sup> Vgl. BORTZ/DÖRING (2002), S. 237; SCHNELL/HILL/ESSER (2005), S. 321; SEIPEL/RIEKER (2003), S. 137.

<sup>6</sup> Vgl. STIER (1996), S. 173.

<sup>7</sup> Vgl. STIER (1996), S. 173.



gung spricht man, wenn lediglich ein bestimmtes Thema vorgegeben ist und die konkrete Gestaltung vollständig den Beteiligten überlassen wird.<sup>1</sup>

Fragen können einerseits nach der Art der Information, die mit ihnen gewonnen werden sollen, und andererseits nach der Struktur der Frage klassifiziert werden. Bezogen auf die *Art der Information* kann zwischen Einstellungs- oder Meinungsfragen, Überzeugungsfragen, Verhaltensfragen und Eigenschaftsfragen differenziert werden.<sup>2</sup> *Einstellungs- oder Meinungsfragen* beziehen sich auf die Wünschbarkeit bzw. die Beurteilung eines Sachverhalts. *Überzeugungsfragen* zielen auf die Wahrnehmung und Einschätzung des Befragten ab und ermitteln, ob der Befragte einen vergangenen, gegenwärtigen oder zukünftigen Sachverhalt für richtig oder falsch bzw. wahr oder unwahr hält. *Verhaltensfragen* beziehen sich auf das tatsächliche oder beabsichtigte Handeln des Befragten, wobei sich dies ebenfalls auf die Vergangenheit, Gegenwart oder Zukunft beziehen kann. *Eigenschaftsfragen* schließlich befassen sich mit demographischen Angaben zum Befragten, wie etwa Alter, Geschlecht, Ausbildung oder Beruf.

Hinsichtlich der Struktur von Fragen ist zwischen offenen und geschlossenen Fragen zu differenzieren.<sup>3</sup> *Offene Fragen* sind dadurch gekennzeichnet, dass sie dem Befragten keine Antwortmöglichkeiten vorgeben und dieser seine Antwort selbst formuliert. Bei *geschlossenen Fragen* hingegen hat der Befragte die Möglichkeit, zwischen vorgegebenen Antworten auszuwählen, wobei zwei oder eine Vielzahl von Alternativantworten vorgegeben werden können. Darüber hinaus kann eine offene und geschlossene Antwortvorgabe *kombiniert* werden (sog. *Hybridfrage*). Hierbei liegt es im Ermessen des Befragten, ob er zusätzlich zu den vorgegebenen Antwortalternativen eine weitere Antwort aufführen möchte.

Nach Art der Durchführung können Befragungen weiterhin in *mündliche Befragungen* (Interview) und *schriftliche Befragungen* (Fragebögen) unterteilt werden.<sup>4</sup> Die Vorteile einer schriftlichen Befragung liegen insbesondere in ihrem geringen Aufwand, da innerhalb eines kürzeren Zeitraums mit weniger Personal eine größere Anzahl an Befrag-

---

<sup>1</sup> Vgl. SEIPEL/RIEKER (2003), S. 149, 153.

<sup>2</sup> Vgl. im Folgenden SCHNELL/HILL/ESSER (2005), S. 325-330 sowie STIER (1996), S. 173-176.

<sup>3</sup> Vgl. im Folgenden ATTESLANDER/KOPP (1995), S. 154-156; SCHNELL/HILL/ESSER (2005), S. 330-334 sowie STIER (1996), S. 176-179.

<sup>4</sup> Vgl. SEIPEL/RIEKER (2003), S. 137. SCHNELL/HILL/ESSER (2005), S. 321 zählen zu den Befragungen weiterhin das *Telefoninterview* sowie die *internetgestützte* Befragung. Bei genauerer Betrachtung können diese jedoch als Sonderform der schriftlichen bzw. mündlichen Befragung identifiziert werden.

ten erreicht werden kann.<sup>1</sup> Hervorzuheben ist weiterhin die Anonymität einer schriftlichen Befragung, was sich positiv auf die Bereitschaft zu ehrlichen Antworten und einer ernsthaften Auseinandersetzung mit der Thematik auswirken kann.<sup>2</sup> Die Nachteile schriftlicher Befragungen sind in der unzureichenden Kontrollierbarkeit der Befragungssituation zu sehen. So können die Antworten der Befragten etwa durch Dritte beeinflusst werden. Weiterhin können Fragen missverstanden werden, da dem Befragten keine Nachfragemöglichkeit zur Verfügung steht. Schließlich besteht die Gefahr, dass Fragen unsorgfältig oder unvollständig bzw. gar nicht beantwortet werden.<sup>3</sup>

Diesen potenziellen Gefahren kann durch die Durchführung einer mündlichen Befragung z.T. entgegengewirkt werden. Der Interviewer kann etwa auf Verständnisfragen eingehen oder bei unvollständigen Antworten „nachhaken“.<sup>4</sup> Insgesamt kann der persönliche Kontakt zu einer höheren Qualität der gewonnenen Daten führen. Allerdings sind mit einer mündlichen Befragung erhebliche Nachteile verbunden. Diese werden in der Literatur auch als „Response-Errors“ bzw. „Antwortverzerrungen“ bezeichnet.<sup>5</sup> Die beiden wichtigsten Formen stellen die Zustimmungstendenz und die soziale Erwünschtheit dar.<sup>6</sup> *Zustimmungstendenz* wird häufig auf die Persönlichkeitsmerkmale von Befragten mit geringer Ich-Stärke zurückgeführt. Sozial erwünschte Antworten hängen zum einen mit dem Bedürfnis nach sozialer Anerkennung des Befragten zusammen und zum anderen mit der Befürchtung von Konsequenzen.

### 3.4.2.2 Beobachtung

Die Beobachtung wird als die „ursprünglichste“ Methode der Datenerhebung angesehen.<sup>7</sup> Dabei wird zu allererst zwischen den Methoden der *direkten* Beobachtung und der *indirekten* Beobachtung differenziert. Erstere bezieht sich auf unmittelbar beobachtbare Eigenschaften oder Handlungen und kann als die Verhaltensbeobachtung im engeren Sinne verstanden werden.<sup>8</sup> Letztere bezeichnet solche Methoden, „die sich ‚nicht auf das Verhalten selbst, sondern auf dessen Spuren, Auswirkungen, Objektivationen‘ be-

---

<sup>1</sup> Vgl. ATTESLANDER (2006), S. 147.

<sup>2</sup> Vgl. BORTZ/DÖRING (2002), S. 237.

<sup>3</sup> Vgl. ATTESLANDER (2006), S. 147.

<sup>4</sup> Vgl. BORTZ/DÖRING (2002), S. 237.

<sup>5</sup> Vgl. SCHNELL/HILL/ESSER (2005), S. 353.

<sup>6</sup> Vgl. im Folgenden SCHNELL/HILL/ESSER (2005), S. 354-355.

<sup>7</sup> Vgl. etwa ATTESLANDER (2006), S. 67.

<sup>8</sup> Vgl. SCHNELL/HILL/ESSER (2005), S. 390.

ziehen“.<sup>1</sup> Die Dokumentenanalyse etwa kann als eine indirekte Beobachtung aufgefasst werden.<sup>2</sup>

Eine der gängigsten Klassifikationen der direkten Beobachtung differenziert zwischen der Feld- und Laborbeobachtung.<sup>3</sup> Unter *Laborbeobachtungen*, mitunter auch als *Experimente* bezeichnet, versteht man die Untersuchung eines Sachverhalts oder eines Vorgangs unter planmäßig simplifizierten, „reinen“ Bedingungen.<sup>4</sup> Sie versuchen „Umweltbedingungen (...) zu simulieren und auf ihrer Grundlage das Verhalten der Objekte in Abhängigkeit der gewählten Stimuli zu erfassen.“<sup>5</sup> Hierbei kommt es entgegen des Wortlauts weniger auf den Ort der Untersuchung an, als auf den Umfang der Kontrolle der Bedingungen, unter denen die Beobachtung durchgeführt wird.<sup>6</sup>

Bei der Durchführung von Laborbeobachtungen sind insbesondere drei Variablen von Bedeutung.<sup>7</sup> Die *unabhängige Variable* ist diejenige Größe, deren Einfluss untersucht werden soll. Als *abhängige Variable* bezeichnet man diejenige Größe, deren Abhängigkeit von der unabhängigen Variable den Gegenstand der Untersuchung darstellt. Als letzte Größe sind die sog. *Störvariablen* zu nennen. Hierunter fallen alle sonstigen Faktoren, die neben den planmäßig variierten unabhängigen Variablen ebenfalls einen Einfluss auf die abhängige Variable haben könnten. Diese Störvariablen sollten im Rahmen des Experiments kontrolliert werden, da ansonsten die Eindeutigkeit der Interpretation gefährdet ist. Zur Kontrolle der Störvariablen existiert eine Vielzahl von Möglichkeiten, etwa die Elimination, die Konstanthaltung, die Parallelisierung, die Umwandlung von Störgrößen in unabhängige Variablen oder die Zufallszuweisung.<sup>8,9</sup>

Die strenge Kontrolle von untersuchungsimmanenten Störvariablen im Rahmen von Laborbeobachtungen gewährleistet, dass Veränderungen der abhängigen Variable mit hoher Wahrscheinlichkeit auf die Variation der unabhängigen Variable zurückgeführt werden kann (sog. interne Validität). Die „Künstlichkeit“ der Untersuchungsumgebung lässt allerdings erhebliche Zweifel daran aufkommen, ob die in einer Stichprobenunter-

---

<sup>1</sup> SCHNELL/HILL/ESSER (2005), S. 391.

<sup>2</sup> Vgl. hierzu ausführlich Kap. 3.4.2.3.

<sup>3</sup> Vgl. etwa BORTZ/DÖRING (2002), S. 60-61, 299 oder HUBER (1995), S. 129-130.

<sup>4</sup> Vgl. ATTESLANDER (2006), S. 168.

<sup>5</sup> BRONNER/WITTE/WODDIDLO (1972), S. 166.

<sup>6</sup> Vgl. BORTZ/DÖRING (2002), S. 299.

<sup>7</sup> Vgl. im Folgenden STELZL (1995), S. 108.

<sup>8</sup> Siehe hierzu ausführlich STAPF (1995), S. 240-241.

<sup>9</sup> BRONNER/WITTE/WODDIDLO (1972), S.166. Allerdings kann diesem Kritikpunkt entgegengehalten werden, dass Experimente bewusst eine Simulation eines Teilausschnitts der Realität anstreben, um eindeutige Aussagen über das Verhalten unter bestimmten und kontrollierten Bedingungen treffen zu können.

suchung gewonnenen Erkenntnisse auch auf andere Personen, Situationen oder Zeitpunkte übertragen bzw. generalisiert werden können (sog. externe Validität).<sup>1</sup>

Im Gegensatz hierzu finden Feldbeobachtungen in einer vom Beobachter ausgewählten, möglichst unbeeinflussten, natürlichen Umgebung statt. Sie repräsentieren daher ein Stück „unverfälschte“ Realität, da das normale Leben ungestört seinen Gang gehen kann.<sup>2</sup> Somit weisen Feldbeobachtungen eine hohe externe Validität auf.<sup>3</sup> Problematisch ist allerdings die geringe interne Validität der Feldbeobachtung, da eine Kontrolle von weiteren in der Realität gegebenen beeinflussenden Variablen (sog. Störvariablen) kaum möglich ist und die Veränderung der abhängigen Variablen häufig auf unterschiedliche Erklärungen zurückgeführt werden kann.<sup>4</sup>

Während Laborexperimente im Rahmen der empirischen Prüfungsforschung eine dominierende Rolle spielen, scheitern Feldbeobachtungen im Bereich des Prüfungswesens bereits deshalb, da die zu erhebenden Daten aufgrund der besonderen Vertrauenssituation zwischen dem Prüfer und Mandanten durch eine hohe Sensibilität gekennzeichnet sind.<sup>5</sup>

### **3.4.2.3 Inhaltsanalyse**

Die Inhaltsanalyse ist eine Methode der Datenerhebung, bei der Informationsträger einer quantifizierenden Analyse unterzogen werden.<sup>6</sup> Als Informationsträger kommen dabei etwa sprachliche Mitteilungen sowie Ton- und Bilddokumente in Betracht. Hauptgegenstand der Inhaltsanalyse stellen jedoch Dokumente dar. Daher ist auch häufig von der Dokumentenanalyse die Rede.<sup>7</sup> Für den Bereich des Prüfungswesens sind insbesondere Jahresabschlüsse, Geschäftsberichte, Bestätigungsvermerke, Prüfungsberichte sowie die Arbeitspapiere einer Abschlussprüfung zu nennen.<sup>1</sup>

Der Inhaltsanalyse liegt der Gedanke zugrunde, dass ein Text „die geronnene Information eines vorausgegangenen Kommunikationsprozesses darstellt, der von einem Bearbeiter nach wissenschaftslogischen Regeln analysiert wird und so Schlußfolgerungen auf

---

<sup>1</sup> Vgl. BORTZ/DÖRING (2002), S. 60, 61 i.V.m. 57.

<sup>2</sup> Vgl. BORTZ/DÖRING (2002), S. 60, 299.

<sup>3</sup> Vgl. BORTZ/DÖRING (2002), S. 60.

<sup>4</sup> Vgl. BORTZ/DÖRING (2002), S. 60.

<sup>5</sup> Vgl. RUHNKE (2000), S. 233-234.

<sup>6</sup> Vgl. SCHNELL/HILL/ESSER (2005), S. 407.

<sup>7</sup> Vgl. STIER (1996), S. 164.

die soziale Wirklichkeit außerhalb des Texts erlaubt.“<sup>2</sup> Die Erschließung der Realität kann sich dabei einerseits auf den Sender sowie den Empfänger eines Texts beziehen, und andererseits auf die soziale Situation, der der Kommunikationsprozess zuzuordnen ist.<sup>3</sup>

Generell lässt sich die Inhaltsanalyse in drei Formen differenzieren.<sup>4</sup> Bei der *Häufigkeitsanalyse* werden Textelemente lediglich klassifiziert und die Häufigkeit ihres Vorkommens erfasst. Im Gegensatz hierzu beschäftigt sich die *Kontingenzanalyse* nicht mit dem Auftreten einzelner Textelemente. Im Mittelpunkt steht vielmehr die Häufigkeit des gemeinsamen Auftretens unterschiedlicher Textelemente, wobei sich für die Darstellung der Zählergebnisse insbesondere die Verwendung von Kreuztabellen – hierbei handelt es sich kombinierte Häufigkeitstabellen für zwei Variablen – eignet. Im Rahmen der *Valenz- und Intensitätsanalysen* erfolgt letztendlich eine Beschreibung eines Texts durch die Abgabe eines Urteils. Im Gegensatz zur Häufigkeits- und Kontingenzanalyse arbeitet man nicht mit nominalskalierten, sondern mit ordinal- oder intervallskalierten Variablen.

Das Hauptproblem der Inhaltsanalyse liegt im Prozess der Datenreduktion, bei der eine Vielzahl von Textelementen in nur wenige Kategorien zu klassifizieren ist. Dabei sollte die Zuverlässigkeit bzw. Konsistenz der Klassifikationen anhand der Kriterien Stabilität, Wiederholbarkeit und Genauigkeit beurteilt werden.<sup>5</sup> Die *Stabilität* einer Inhaltsanalyse setzt voraus, dass die Inhaltsklassifikation durch die gleiche Person zu unterschiedlichen Zeitpunkten zum identischen Ergebnis (sog. Intracoder-Reliabilität) führt. Die *Wiederholbarkeit* zielt hingegen auf die Vergleichbarkeit von Ergebnissen ab, wenn die Analyse des identischen Texts durch unterschiedliche Personen zum gleichen Zeitpunkt durchgeführt wird (sog. Intercoder-Reliabilität). Die *Genauigkeit* schließlich bezieht sich auf die Übereinstimmung der Klassifikationen einer bekannten Standardcodierung bzw. Codierungsnorm.

---

<sup>1</sup> Vgl. RUHNKE (2000), S. 234 Fn. 8 m.w.N.

<sup>2</sup> ATTESLANDER (1993), S. 221.

<sup>3</sup> Vgl. ATTESLANDER (2006), S. 182-183.

<sup>4</sup> Vgl. im Folgenden BORTZ/DÖRING (2002), S. 151.

<sup>5</sup> Vgl. im Folgenden SCHNELL/HILL/ESSER (2005), S. 412-413.

### 3.4.3 Methoden der Datenauswertung

Nachdem die Daten erhoben und aufbereitet wurden, müssen sie in einem nächsten Schritt ausgewertet bzw. analysiert werden. Hierzu stehen unterschiedliche statistische Verfahren zur Verfügung. Generell kann hierbei zwischen der *deskriptiven Statistik*, welche der statistischen Beschreibung der erhobenen Daten dient, und der *schließenden* bzw. *Inferenzstatistik*, auf die etwa zur Überprüfung von Hypothesen zurückgegriffen wird, differenziert werden.<sup>1</sup> Dabei kommen univariate, bivariate und multivariate Auswertungsverfahren in Betracht.

*Univariate Verfahren* werden insbesondere zur Charakterisierung der Verteilung bestimmter Merkmale in einer Stichprobe eingesetzt.<sup>2</sup> Hierzu zählen neben der Darstellung von Häufigkeiten, etwa anhand von Histogrammen bzw. Balken- und Kreisdiagrammen, insbesondere die Verwendung von Lage- und Streuungsmaßen. Als Lagemaße kommen der Modus, der Median und das arithmetische Mittel in Betracht, welche Aussagen über typische Werte einer Verteilung ermöglichen. Streuungsmaße, wie etwa Spannweite, Standardabweichung und Varianz, informieren hingegen über das Ausmaß von Unterschiedlichkeiten der erhobenen Daten.<sup>3</sup>

Möchte man hingegen die Zusammenhänge von zwei Variablen untersuchen, so ist der Einsatz von *bivariablen Verfahren* notwendig. Diese ermöglichen die Untersuchung von theoretischen Fragestellungen, indem Hypothesen oder Forschungsfragen anhand von Signifikanztests systematisch beurteilt werden.<sup>4</sup> Das einfachste Verfahren stellt dabei der *Chi-Quadrat-Verteilungstest* dar, welcher der Überprüfung von Hypothesen über bestimmte Verteilungsannahmen dient.<sup>5, 6</sup> Mit Hilfe dieses Testverfahrens kann beurteilt werden, ob Abweichungen zwischen erwarteten und beobachteten Werten signifikant sind.

Bei der Analyse von Daten ist regelmäßig von Interesse, ob Unterschiede zwischen Gruppen zufällig sind oder eher auf systematische Ursachen zurückgeführt werden können. Hierzu können Signifikanztests herangezogen werden, die etwa die Differenzen in

---

<sup>1</sup> Vgl. RUHNKE (2000), S. 235.

<sup>2</sup> Vgl. SEIPEL/RIEKER (2003), S. 178.

<sup>3</sup> Vgl. ATTESLANDER (2006), S. 246-249.

<sup>4</sup> Vgl. SEIPEL/RIEKER (2003), S. 179.

<sup>5</sup> Vgl. ATTESLANDER (2006), S. 268.

<sup>6</sup> Zu differenzieren ist der *Chi-Quadrat-Verteilungstest* insbesondere von dem *Chi-Quadrat-Unabhängigkeitstest*. Während der Verteilungstest ein parametrischer Test ist, handelt es sich beim Unabhängigkeitstest um ein nicht-parametrisches Testverfahren. Vgl. hierzu auch die nachfolgenden Ausführungen.

den Mittelwerten unterschiedlicher Gruppen überprüfen.<sup>1</sup> Zu differenzieren ist hierbei generell zwischen parametrischen und nicht-parametrischen Testverfahren. *Parametrische Testverfahren* weisen hohe Anforderungen an die Beschaffenheit der zu überprüfenden Daten auf, wie etwa das Vorliegen einer Normalverteilung oder die Gleichheit der Varianzen.<sup>2</sup> Hierzu zählen der *t-Test* und die *einfaktorielle Varianzanalyse*. Während beim *t-Test* untersucht wird, ob sich die Mittelwerte zweier Gruppen signifikant voneinander unterscheiden, werden bei der einfaktoriellen Varianzanalyse die Mittelwerte von mehr als zwei Gruppen miteinander verglichen.<sup>3</sup> Als die nicht-parametrischen Pendanten dieser beiden Testverfahren gelten der *Mann-Whitney-U-Test* bei zwei Gruppen und der *Kruskal-Wallis-Test* bei mehr als zwei Gruppen. Ein weiterer nicht-parametrischer Test ist in dem *Chi-Quadrat-Unabhängigkeitstest* zu sehen, welcher überprüft, ob zwei Merkmale statistisch unabhängig voneinander sind. Nicht-parametrische Testverfahren werden häufig dann eingesetzt, wenn die zu beurteilenden Daten nicht den Anforderungen der parametrischen Tests genügen.<sup>4, 5</sup>

In vielen empirischen Untersuchungen sind jedoch *multivariate Verfahren* erforderlich, um die Komplexität der Zusammenhänge angemessen berücksichtigen zu können. Hierbei handelt es sich um Analysen, die den Einfluss einer Vielzahl von Faktoren simultan überprüfen. Hierzu gehören die Regressions- und die Varianzanalyse. Diese beiden Verfahren ermöglichen es, den Einfluss mehrerer unabhängiger Faktoren auf die abhängige Variable zu untersuchen und eignen sich insbesondere zur Überprüfung komplexerer Hypothesensysteme.<sup>6</sup>

---

<sup>1</sup> Vgl. SCHNELL/HILL/ESSER (2005), S. 451.

<sup>2</sup> Vgl. BORTZ (2005), S. 141, 284.

<sup>3</sup> Vgl. SEIPEL/RIEKER (2003), S. 182-183.

<sup>4</sup> Vgl. BORTZ (2005), S. 150, 287.

<sup>5</sup> Ausführliche Darstellungen zu nichtparametrischen statistischen Methoden finden sich bei SIEGEL (1997).

<sup>6</sup> Vgl. SEIPEL/RIEKER (2003), S. 183.

### **3.5 Darstellung und Analyse der empirischen Prüfungsforschung zur Durchführung von analytischen Prüfungshandlungen vor dem Hintergrund des abgeleiteten Problemlösungsmodells**

#### **3.5.1 Vorbemerkung**

Im Folgenden sollen die umfassenden und überwiegend aus dem angloamerikanischen Sprachraum hervorgegangenen Ergebnisse anhand des unter Kap. 3.3 deduzierten Problemlösungsmodells systematisiert werden. Die Studienergebnisse erfahren dabei eine kurze Darstellung. Da es sich zumeist um Experimente handelt, soll kurz auf den Versuchsaufbau eingegangen werden, um anschließend auf die aufgestellten Forschungshypothesen – hierin spiegelt sich *Schritt 3* des deduzierten verhaltensorientierten Forschungsprogramms wider – der jeweiligen Studien einzugehen. Darüber hinaus sollen die Studienergebnisse – welche durch Beobachtung des Experiments (*Schritt 4* des Forschungsprogramms) und Auswertung der erhobenen Daten (*Schritt 5* des Forschungsprogramms) gewonnen wurden – darauf hin untersucht werden, ob sie sich im Sinne einer technologischen Aussage zur Ableitung von Handlungsempfehlungen für die Prüfungspraxis eignen.

#### **3.5.2 Studien zur Phase der Problemrepräsentation**

##### **3.5.2.1 Aufbau einer adäquaten Problemrepräsentation und der Einfluss von Wissensstrukturen auf das Problemverständnis**

In der psychologischen Forschung ist man überwiegend der Auffassung, dass die Quantität von Wissen und dessen Strukturierung einen wesentlichen Einfluss auf die Entwicklung einer adäquaten Problemrepräsentation ausübt.<sup>1</sup> Individuen, die über umfangreiches und gut strukturiertes Wissen verfügen, sind in der Lage, komplexe Informationsmuster zu erkennen und somit einen umfassenden und gut strukturierten Problemraum aufzubauen sowie eine für die Bearbeitung des Problems angemessene Problemlösungsstrategie auszuwählen.<sup>2, 3</sup>

---

<sup>1</sup> Vgl. CHASE/SIMON (1973), S. 215-281; MCKEITHEN/REITMAN/RUETER/HIRTLE (1981), S. 307-325; LARKIN (1983), S. 75-99.

<sup>2</sup> Vgl. LESGOLD/RUBINSON/FELTOVICH/GLASER/KLOPFER/WANG (1988), S. 311-342.

<sup>3</sup> In der experimentellen Prüfungsforschung versucht man Wissensstrukturen anhand von Sortieraufgaben bzw. Erinnerungsaufgaben zu identifizieren. Bei Sortieraufgaben sollen Probanden mit unterschiedlichen Wissensständen, und somit auch mit Unterschieden in der Menge der vorhandenen Wis-



Eine der ersten Studien, die sich auf die Problemrepräsentation von Prüfern bei der Durchführung von analytischen Prüfungshandlungen bezieht, wurde von BIGGS/MOCK/WATKINS durchgeführt. Sie ließen insgesamt vier Prüfer (zwei Manager<sup>1</sup> mit jeweils sieben Jahren Praxiserfahrung und zwei Prüfungsleiter<sup>2</sup> mit jeweils weniger als drei Jahren Praxiserfahrung) einen Prüfungsplan erstellen, wobei insbesondere auf die Ergebnisse von analytischen Prüfungshandlungen abzustellen war.<sup>3</sup> Die Fallstudie umfasst eine Überbewertung des Forderungsbestands, die sich darin niederschlug, dass der Forderungszuwachs wesentlich höher ausfiel als die Erhöhung der Abschreibungen auf Forderungen einerseits und der zweifelhaften Forderungen andererseits.<sup>4</sup> Da es sich hier um einen Fehler handelte, der durch einfache analytische Prüfungshandlungen aufgedeckt werden konnte, ist es wenig verwunderlich, dass alle Prüfer in der Lage waren, die unerwartete Abweichung zu identifizieren. Dennoch konnten BIGGS/MOCK/WATKINS erhebliche Unterschiede in der Vorgehensweise der Prüfer feststellen. Die Manager benötigten im Gegensatz zu den Prüfungsleitern wesentlich weniger Zeit, um ein umfassendes Problemverständnis für den zu prüfenden Sachverhalt aufzubauen. Die Prüfungsleiter waren einen großen Teil ihrer Zeit damit beschäftigt, sich erneut in die Fallbeschreibung einzulesen. Die Autoren führten dieses Verhalten auf das Fehlen von entsprechenden Wissensstrukturen zurück. Während die Manager zügig eine Problemrepräsentation entwickeln konnten, die es ihnen erleichterte, die für eine erfolgreiche Problemlösung notwendigen Informationen zu erkennen, beschäftigten sich die Prüfungsleiter auch mit irrelevanten Informationen der Fallbeschreibung.<sup>1</sup>

Weiterhin konnte festgestellt werden, dass die entwickelte Problemrepräsentation der Prüfungsleiter im Wesentlichen auf die oberflächlichen Informationen der Aufgabenstellung gerichtet war. So betrachteten sie lediglich Informationen, die es ihnen ermöglichten, die notwendigen Prüfungsschritte durchzuführen. Die Manager konzentrierten

---

sensstrukturen, unterschiedliche Sachverhalte anhand eines selbst gewählten Merkmals klassifizieren. Eine prüfungsbezogene Studie hierzu führten FREDERICK/HEIMAN-HOFFMAN/LIBBY (1994), S. 7-12 durch. Einen weitaus stärkeren Anklang im Bereich der Prüfungsforschung finden jedoch Erinnerungsaufgaben. Hierbei sollen sich Probanden mit unterschiedlichen Wissensständen an möglichst viele Aspekte eines Sachverhalts erinnern, die ihnen zuvor dargeboten wurde. Prüfungsbezogene Studien hierzu finden sich bei WEBER (1980), S. 217-223, FREDERICK (1991), S. 251, BHATTACHARYA (1994), S. 62-64, 93-94 sowie CHRIST (1993), S. 308-318.

<sup>1</sup> Der Manager stellt auf der internen Mitarbeiterhierarchie einer Prüfungsgesellschaft die zweithöchste Stufe dar. Er ist gegenüber dem Partner weisungsgebunden und gegenüber dem Prüfungsleiter weisungsberechtigt. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 524.

<sup>2</sup> Der Prüfungsleiter ist für den ordnungsgemäßen Ablauf der Prüfung beim Mandanten verantwortlich. Er ist auf der internen Mitarbeiterhierarchie einer Prüfungsgesellschaft oberhalb der Prüfungsassistenten und unterhalb der Partner und Manager. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 622.

<sup>3</sup> Vgl. BIGGS/MOCK/WATKINS (1988), S. 150.

<sup>4</sup> Vgl. BIGGS/MOCK/WATKINS (1988), S. 155-156.

sich darüber hinaus auch auf Informationen, die Rückschlüsse auf den Mandanten, seinen Geschäftsbetrieb und seine Geschäftsprobleme zuließen.<sup>2</sup> Dieses lässt darauf schließen, dass Manager über mehr oder zumindest qualitativ hochwertigere Wissensstrukturen verfügen, die es ihnen erlauben, ein umfassendes Problemverständnis über den zugrunde liegenden Sachverhalt zu entwickeln.

Diese These wird auch durch einen Vergleich der weiteren Vorgehensweisen gestützt.<sup>3</sup> Wie zuvor erwähnt, sind alle Prüfer auf das erhöhte Risiko im Forderungsbestand aufmerksam geworden. Die Prüfungsentscheidungen der Prüfungsleiter offenbarten jedoch Fehleinschätzungen bezüglich der weiteren vorzunehmenden Prüfungshandlungen, mit denen letztendlich die Werthaltigkeit der ausgewiesenen Forderungen überprüft werden soll. Die Prüfungsleiter erhöhten eine Vielzahl von Prüfungshandlungen, darunter auch solche, die nicht auf das eigentliche Ziel, die Werthaltigkeitsüberprüfung, ausgerichtet waren, sondern auf diverse andere Prüfungsaussagen.<sup>4</sup> Manager hingegen gingen weit- aus selektiver vor und erhöhten lediglich jene Prüfungshandlungen, die auf die fragliche Werthaltigkeit der Forderungen ausgerichtet waren. BIGGS/MOCK/WATKINS führen die überlegene Vorgehensweise der Manager auf ihre umfassendere Problemrepräsentation zurück, die es ihnen erlaubt, ein tiefergehendes und strukturierteres Verständnis für den betrachteten Sachverhalt aufzubringen.

Zusammenfassend bleibt festzuhalten, dass BIGGS/MOCK/WATKINS den positiven Einfluss von Wissensstrukturen auf die Problemrepräsentation und somit auf die gesamte Problemlösung bestätigten. Zwar konnte gezeigt werden, dass auch weniger erfahrene Prüfer (Prüfungsleiter) in der Lage sind, die unerwartete Abweichung zu identifizieren. Dennoch benötigen sie für die Entwicklung eines entsprechenden Problemverständnisses wesentlich mehr Zeit als erfahrenere Prüfer (Manager). Somit kann geschlussfolgert werden, dass adäquate Wissensstrukturen wesentliche Effizienzvorteile mit sich bringen. Darüber hinaus konnte festgestellt werden, dass das mangelnde Problemverständnis der unerfahrenen Prüfer (Prüfungsleiter) als Ursache dafür angesehen werden kann, dass Fehlentscheidungen bezüglich der Ausweitung von Prüfungshandlungen getroffen wurden. Die hierdurch bestehende Gefahr einer Nicht-Aufdeckung eines wesentlichen

---

<sup>1</sup> Vgl. BIGGS/MOCK/WATKINS (1988), S. 153.

<sup>2</sup> Vgl. BIGGS/MOCK/WATKINS (1988), S. 153.

<sup>3</sup> Vgl. im Folgenden BIGGS/MOCK/WATKINS (1988), S. 157.

<sup>4</sup> Im Zentrum des Interesses liegt hier die Ausweitung der Saldenbestätigungen. Da eine Analyse der Altersstrukturlisten der Forderungen auf eine Veralterung des Forderungsbestands hindeutete, käme es vielmehr darauf an, die Werthaltigkeit der ausgewiesenen Forderungen zu überprüfen. Saldenbestäti-

Fehlers führt zu erheblichen Effektivitätseinbußen. Darüber hinaus erhöhen Prüfungshandlungen, die nicht zur Aufdeckung der eigentlichen Ursache geeignet sind, die weiteren Prüfungskosten, was wiederum die Effizienz verringert. Es bleibt demnach festzuhalten, dass analytische Prüfungshandlungen im Rahmen der Prüfungsplanung nach Möglichkeit von erfahreneren Prüfern, also Managern, durchgeführt werden sollten, die über ausreichende und qualitativ hochwertige Wissensstrukturen verfügen, da ansonsten mit erheblichen Effektivitäts- und Effizienzeinbußen zu rechnen ist. Zum Teil steht dieses allerdings im Widerspruch zur Prüfungspraxis, da die Prüfungsplanung regelmäßig von Prüfungsleitern durchgeführt wird.<sup>1</sup>

Eine weitere Studie, die sich mit der Problemrepräsentation von Prüfern bei der Durchführung von analytischen Prüfungshandlungen beschäftigt, führten BEDARD/BIGGS durch. Sie ließen insgesamt 21 Prüfer (zehn Manager und elf Prüfungsleiter<sup>2</sup>) eine Fallstudie bearbeiten, die mit einem Fehler bei der Gemeinkostenallokation im Bereich der Vorräte versehen war. Als Resultat konnte ein im Vergleich zum Vorjahr erhöhter Vorratsbestand sowie erhöhter Gewinn bei gleich bleibenden Umsätzen identifiziert werden, was auf eine Überbewertung der Vorräte und des Gewinns schließen lässt.<sup>3</sup> Im Gegensatz zur BIGGS/MOCK/WATKINS-Studie wurde den Probanden nicht die Erstellung eines Prüfungsplans zur Aufgabe gestellt, sondern die Identifikation der Ursache für die festgestellte Abweichung.

BEDARD/BIGGS konzipierten ihre Fallstudie derart, dass es durchaus möglich war, den Grund für die unerwartete Abweichung allein durch Anwendung analytischer Prüfungshandlungen, also ohne weitere Einzelfallprüfungen, zu identifizieren.<sup>4</sup> Entscheidend für die Problemlösung war, dass die Ursache für den Fehler nicht festgestellt werden konnte, bevor der Prüfer die einzelnen Abweichungen als ein zusammenhängendes Muster erkannt hat.<sup>5</sup> Die Mustererkennung kann als ein zentraler Aspekt sowohl der regel- als auch der schemabasierten Problemrepräsentation bezeichnet werden.<sup>6</sup>

---

gungen hingegen zielen vielmehr auf die Überprüfung der Existenz und der Richtigkeit ab. Vgl. BIGGS/MOCK/WATKINS (1988), S. 157.

<sup>1</sup> Vgl. HIRST/KOONCE (1996), S. 462.

<sup>2</sup> Dabei wurden keine Angaben bezüglich der Berufserfahrung der Prüfer gemacht.

<sup>3</sup> Vgl. BEDARD/BIGGS (1991a), S. 626-627.

<sup>4</sup> Einer unerwarteten Erhöhung der Vorräte um 150.000 \$ stand eine unerwartete Erhöhung des Gewinns nach Steuern um 90.000 \$ gegenüber. Da von einem Steuersatz von 40% auszugehen war, belief sich die unerwartete Gewinnerhöhung auf brutto 150.000 \$. Folglich konnten die Prüfer auf eine fehlerhafte Aktivierung der (ergebniswirksamen) Gemeinkosten i.H.v. 150.000 \$ schließen. Vgl. BEDARD/BIGGS (1991a), S. 627, 630-631.

<sup>5</sup> Vgl. BEDARD/BIGGS (1991a), S. 626.

<sup>6</sup> Vgl. Hierzu ausführlich Kap.3.2.2.3.4.2.2.3 sowie 3.2.2.3.4.2.2.4.

Von den 21 Prüfern gelang es lediglich sechs Prüfern (fünf Managern und einem Prüfungsleiter) die richtige Ursache für den Fehler festzustellen. Prüfer, welche die Aufgabe erfolgreich bearbeitet haben, waren in der Lage, alle festgestellten Abweichungen in ein passendes Muster zu integrieren und so das relevante Wissen über die Ursache des Musters abzurufen.<sup>1</sup> Die Vorgehensweise von nicht erfolgreichen Prüfern zeichnete sich u.a. dadurch aus, dass lediglich ein bis zwei Hinweise in die Problemrepräsentation aufgenommen wurden.<sup>2</sup> Da es sich nach BEDARD/BIGGS bei dem zur Lösung des betrachteten Problems notwendigen Wissen um grundlegende rechnungslegungspezifische Kenntnisse (insbesondere Kenntnisse über die doppelte Buchführung und die Gemeinkostenallokation) handelte, von denen zu vermuten war, dass sie den teilnehmenden Prüfern generell bekannt waren, gingen die Autoren davon aus, dass sich die nicht erfolgreichen Prüfer lediglich auf die „oberflächlichen“ Hinweise der Fallstudie konzentrierten. Demzufolge waren sie nicht im Stande, diese Hinweise miteinander zu kombinieren, was ein tiefergehendes Wissen über Jahresabschlusspositionen bzw. hierauf basierende Kennzahlen und deren Relationen zueinander voraussetzt.<sup>3</sup> BEDARD/BIGGS schlussfolgerten daraus, dass die mangelnde Fähigkeit zur Erfassung von Hinweisen in einem Muster (sog. Mustererkennung) darauf zurückzuführen sei, dass analytische Prüfungshandlungen in der Prüfungspraxis häufig nur in Form von einfacheren Methoden, etwa einfachen Vorjahresvergleichen, durchgeführt werden und weniger auf Kombination von multiplen Hinweisen ausgerichtet sind.<sup>4,5</sup>

Die Schlussfolgerungen von BEDARD/BIGGS weisen auf einen wesentlichen Aspekt hin, der mit den Erkenntnissen der bisherigen Untersuchung folgendermaßen in Einklang gebracht werden kann. Der Grund für die mangelnde Fähigkeit zur Kombination von Daten bzw. zur Mustererkennung könnte darin liegen, dass derartige Prüfungshandlungen in der Praxis kaum durchgeführt werden. Die fehlende Erfahrung der Prüfer verhindert die Entwicklung von entsprechenden Wissensstrukturen, d.h. das Wissen ist noch nicht ausreichend vernetzt. Mit anderen Worten bedeutet dies, dass die Prüfer zwar über

---

<sup>1</sup> Vgl. BEDARD/BIGGS (1991a), S. 632.

<sup>2</sup> Vgl. BEDARD/BIGGS (1991a), S. 635.

<sup>3</sup> Vgl. BEDARD/BIGGS (1991a), S. 635-636.

<sup>4</sup> Vgl. BEDARD/BIGGS (1991a), S. 636. Eine überwiegende Verwendung einfacherer Methoden im Rahmen der Prüfungsplanung sowie zum Teil auch während der Prüfungsdurchführung konnte in diversen weiteren angloamerikanischen Studien nachgewiesen werden. Vgl. etwa HIRST/KOONCE (1996), S. 470 sowie FRASER/HATHERLY/LIN (1997), S. 41-42 und LIN/FRASER (2003), S. 159-160.

<sup>5</sup> Für den nationalen Bereich stellte GÄRTNER (1994), S. 180-181 ebenfalls eine verstärkte Anwendung von einfacheren Methoden fest. In einer dem Verfasser vorliegenden, allerdings unveröffentlichten empirischen Untersuchung aus dem Jahr 2002 konnte dieser Sachverhalt wiederholt festgestellt werden.

das entsprechende Wissen verfügen (sog. Faktenwissen), dieses Wissen aber „träge“ und nur schwierig im Rahmen einer Problemrepräsentation abzurufen ist. Erst durch wiederholte Anwendung derartiger Prüfungshandlungen können die einzelnen Wissensseinheiten miteinander verbunden werden, was ihre Abrufbarkeit für die spätere Problemrepräsentation erheblich erleichtert. Es ist also zu vermuten, dass den nicht erfolgreichen Prüfern keine entsprechenden Wissensstrukturen zur Verfügung standen und sie im Rahmen der Problemrepräsentation nur die oberflächlichen Hinweise verarbeiten konnten.

Für eine derartige Vermutung spricht auch die Studie von BIERSTAKER/BEDARD/BIGGS, die zwölf Prüfungsleiter die gleiche Fallstudie bearbeiten ließen wie BEDARD/BIGGS (sog. Replikationsstudie).<sup>1</sup> Die Autoren wählten in ihrer Studie ausschließlich Prüfungsleiter, da sich bei BEDARD/BIGGS zeigte, dass es gerade diese Gruppe von Prüfern war, die nicht auf das für die Bearbeitung der Aufgabe benötigte Wissen zugreifen konnten.<sup>2</sup> BIERSTAKER/BEDARD/BIGGS konnten die Ergebnisse von BEDARD/BIGGS bestätigen. Von den zwölf Probanden gelang es lediglich einem, die Aufgabe ohne Hilfestellung erfolgreich zu bearbeiten.<sup>3</sup> Im Unterschied zu BEDARD/BIGGS ließen BIERSTAKER/BEDARD/BIGGS den Probanden nach den anfänglichen Misserfolgen sog. „prompts“ zukommen, die ihnen einen Hinweis auf den richtigen Lösungsansatz geben und somit das Abrufen des für die Problemlösung notwendigen Wissens erleichtern sollten.<sup>4</sup> Dabei waren die Autoren bemüht, die prompts so realitätsnah wie möglich zu konzipieren („practical plausibility“), etwa als ein von der eingesetzten Prüfungssoftware generierter Hinweis im Rahmen der Prüfungsdokumentation.<sup>5</sup> Nach Mitteilung der prompts gelang es zehn weiteren Prüfern, die Aufgabe erfolgreich abzuschließen.<sup>6</sup> Da es sich bei den prompts um „Denkanstöße“ handelte, von denen angenommen werden kann, dass deren Inhalte sicherlich nicht das Wissen der Prüfungsleiter übersteigt, liegt die Vermutung nahe, dass die Prüfungsleiter ihr Wissen nicht entsprechend abrufen konnten. Es kann davon ausgegangen werden, dass die fehlende Strukturierung des Wissens für die „Trägheit“ des vorhandenen Wissens und somit für die oberflächliche

---

<sup>1</sup> Da im Mittelpunkt der Studie von BIERSTAKER/BEDARD/BIGGS allerdings nicht der Einfluss von Wissensstrukturen auf die Problemrepräsentation stand, sondern der Wechsel von einer „unproduktiven“ zu einer „produktiven“ Problemrepräsentation sowie Faktoren, die einen derartigen Wechsel begünstigen bzw. beeinträchtigen, erfolgt eine ausführliche Diskussion dieser Studie erst unter Kap. 3.5.2.3.

<sup>2</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 26.

<sup>3</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 27 i.V.m. S. 29.

<sup>4</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 23.

<sup>5</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 23-24.

<sup>6</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 29.

Problembearbeitung verantwortlich war. Erst durch externe Hinweise gelang es den Probanden, ihr Wissen zu repräsentieren und ein tieferes Verständnis für den zugrunde liegenden Sachverhalt aufzubauen, was in einer erfolgreichen Problemlösung resultierte.

Berücksichtigt man die Ergebnisse der beiden vorangestellten Studien, so ergeben sich für die Prüfungspraxis folgende Implikationen: Analytische Prüfungshandlungen, bei denen Daten kombiniert werden müssen, um ein auf die Abweichungsursache hinweisendes Muster zu erkennen, werden in der Prüfungspraxis selten angewandt. Es dominieren einfachere Methoden, insbesondere Vorjahresvergleiche mit einzelnen Jahresabschlusspositionen oder Kennzahlen. Dieses führt dazu, dass die Prüfungsleiter bei der Durchführung von analytischen Prüfungshandlungen nicht in der Lage sind, multiple Abweichungen in den prüfungsrelevanten Daten in ein Muster zu integrieren, was sie davon abhält, den zugrunde liegenden Fehler zu identifizieren. Lediglich Manager, also erfahrene Prüfer, waren ohne Hilfestellung imstande, die Muster im Prüffeld zu erkennen und somit eine effektive und effiziente Prüfungsdurchführung voranzutreiben. Es ist also davon auszugehen, dass erst nach vieljähriger Prüfungserfahrung, wie es etwa bei Managern der Fall ist, umfassende Wissensstrukturen aufgebaut werden.

Da jedoch der Hauptanteil der Prüfungstätigkeit im Rahmen der Durchführungsphase nicht von den Managern durchgeführt wird<sup>1</sup>, lassen die Ergebnisse von BEDARD/BIGGS schwerwiegende Mängel in der Prüfungspraxis vermuten. Zwar hätte die Fehlerursache auch durch andere Prüfungsaktivitäten, etwa durch eine Rückermittlung des Inventurbestands oder zusätzliche compliance tests, identifiziert werden können<sup>2</sup>, womit nicht per se von Effektivitätseinbußen ausgegangen werden kann. Da der Grund für die Überbewertung jedoch allein durch Kombination von mehreren Hinweisen festzustellen war, hätten derartige kostenintensive Prüfungsaktivitäten durch die Anwendung einer analytischen Prüfungshandlung vermieden werden können, so dass hier zumindest von nicht unwesentlichen Effizienzeinbußen ausgegangen werden kann. Es ist also zu überlegen, welche Maßnahmen ergriffen werden können, um derartige Ineffizienzen zu vermeiden.

BIERSTAKER/BEDARD/BIGGS zeigten in ihrer Studie, dass der mangelnden Fähigkeit der Prüfungsleiter, ihr Wissen im Rahmen der Problemlösung abzurufen, durch das Mitteilen von „Denkanstößen“ entgegen gewirkt werden konnte. Folglich könnte darüber nachgedacht werden, dass derartige Hilfestellungen im Rahmen der Prüfungsdokumen-

---

<sup>1</sup> Vgl. KOONCE/PHILLIPS (1996), S. 35-36 m.w.N. sowie ANDERSON/KADOUS/KOONCE (2004), S. 15 m.w.N.

<sup>2</sup> Vgl. BEDARD/BIGGS (1991a), S. 626.

tation von der jeweils eingesetzten Prüfungssoftware aufgerufen werden. Software zur Unterstützung der Prüfungsdokumentation genießt in der heutigen Prüfungspraxis eine hohe Akzeptanz. Demnach wäre die Umsetzung dieser Maßnahme ohne hohen Aufwand durchaus denkbar. Da hierdurch die Möglichkeit bestünde, die Wirtschaftlichkeit der einzelnen Prüfungsaufträge zu erhöhen, wäre der verhältnismäßig geringe Aufwand sicherlich lohnenswert.

Eine weitere adäquate Maßnahme könnte darin zu sehen sein, Prüfer im Rahmen ihrer Aus- und Weiterbildung frühzeitiger mit der Anwendung von analytischen Prüfungshandlungen, bei denen unterschiedliche Daten zu einem Muster zusammengefügt werden müssen, vertraut zu machen. Hierfür bieten sich insbesondere Fallstudien an, die sowohl im Rahmen der universitären Ausbildung mit dem Schwerpunkt Prüfungswesen als auch in Fortbildungsseminaren der Prüfungsgesellschaften aufgenommen werden könnten. Eine weitere Möglichkeit könnte auch in der Aufnahme derartiger Beispiele in die Prüfungsliteratur darstellen.

### **3.5.2.2 Verzerrungen beim Aufbau der Problemrepräsentation**

#### **3.5.2.2.1 Framing-Effekte**

Je nachdem, aus welcher Perspektive man einen Problemsachverhalt betrachtet, kann sich beim Problemlöser eine unterschiedliche Repräsentation des Problemsachverhalts entwickeln und somit die gesamte Problemlösung maßgeblich beeinflusst werden.<sup>1</sup> TVERSKY/KAHNEMAN weisen darauf hin, dass verschiedene Perspektiven allein durch die sprachliche Formulierung in der Aufgaben- bzw. Fragestellung verursacht werden können.<sup>2</sup> Derartige ausschließlich durch die Wortwahl beeinflusste Problemrepräsentationen werden in der Literatur auch als *Framing-Effekte* bezeichnet.<sup>3</sup>

Da sowohl die berufsrechtlichen Regelungen als auch die Prüfungshandbücher der einzelnen Wirtschaftsprüfungsgesellschaften in Bezug auf die verschiedenen Prüfungsaufgaben unterschiedliche Formulierungen enthalten, hat sich die Prüfungsforschung in den letzten beiden Jahrzehnten intensiv mit dem Einfluss von Framing-Effekten auf die

---

<sup>1</sup> Vgl. EMDY/FINNLEY (1997), S. 63; JAMAL/JOHNSON/BERRYMAN (1994), S. 86.

<sup>2</sup> Vgl. TVERSKY/KAHNEMAN (1981), S. 453-458; TVERSKY/KAHNEMAN (1986), S. 257; siehe auch KAHNEMAN/TVERSKY (1984), S. 341-350.

<sup>3</sup> Vgl. etwa TROTMAN/SNG (1989), S. 565; AYERS/KAPLAN (1993), S. 113-114; EMBDY (1994), S. 103-104; CHANG/YEN/DUH (2002), S. 36.

Problemrepräsentation beschäftigt.<sup>1</sup> Im Mittelpunkt der Forschungsbemühungen standen zunächst die Untersuchungen zu Going-Concern-Beurteilungen.<sup>2</sup> Allerdings konnten hier keine signifikanten Einflüsse von Framing auf die Problemrepräsentation festgestellt werden. Diese Ergebnisse könnten jedoch auf zwei wesentliche Gründe zurückgeführt werden. Erstens sind Going-Concern-Beurteilungen für den Prüfer mit einem immens hohen Risiko verbunden. Daher könnten Prüfer äußerst sensibel auf Informationen reagieren, die auf eine Gefährdung der Unternehmensfortführung deuten.<sup>3</sup> Eine derartige konservative Strategie wird in der Literatur häufig auch als *professionelle Skepsis* bezeichnet, die möglicherweise dazu führen könnte, dass Framing-Effekte bei Going-Concern-Beurteilungen überlagert werden.<sup>4</sup> Zweitens führten sowohl KIDA als auch TROTMAN/SNG ihre Studien mit Prüfern durch, die sich durch ein relativ hohes Maß an Berufserfahrung auszeichneten.<sup>5</sup> Diese Vorgehensweise kann zwar als durchaus realistisch bezeichnet werden, da es überwiegend erfahrene Prüfer sind, die in der Prüfungspraxis für Going-Concern-Entscheidungen verantwortlich sind. Für die Untersuchung von Framing-Effekten scheinen sie jedoch aufgrund folgender Überlegung als weniger zweckdienlich: erfahrene Prüfer zeichnen sich gegenüber unerfahrenen Prüfern dadurch aus, dass sie über eine höhere Anzahl von stärker ausgeprägten Wissensstrukturen verfügen. CHURCH vermutet, dass sie hierdurch in der Lage seien, den Problem Sachverhalt objektiver und somit weniger abhängig von der jeweiligen Wortwahl repräsentieren zu können. Demzufolge wären erfahrene Prüfer tendenziell weniger von Framing-Effekten im Rahmen der Problemrepräsentation betroffen.<sup>6</sup> Da jedoch im Rahmen einer Abschlussprüfung auch unerfahrene Prüfer zum Einsatz kommen, kann festgehalten werden, dass die o.a. Untersuchungen von Framing-Effekten nur teilweise herangezogen werden können.<sup>7</sup>

Ein Prüfungsbereich, der weitaus geeigneter als Going-Concern-Beurteilungen sein könnte, um Framing-Effekte zu untersuchen, ist in der Durchführung von analytischen

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 121.

<sup>2</sup> Vgl. etwa KIDA (1984), S. 332-340; TROTMAN/SNG (1989), S. 565-576; CLOCK/DEVINE (1995), S. 197-207.

<sup>3</sup> Vgl. SCHREIBER (2000), S. 123-124 m.w.N.

<sup>4</sup> Vgl. etwa SCHREIBER (2000), S. 122 m.w.N.

<sup>5</sup> An der Studie von KIDA nahmen ausschließlich Partner und Manager teil, die in der Regel eine Berufserfahrung von 5-6 Jahren aufweisen. Die Prüfer, die an der Studie von TROTMAN/SNG teilnahmen, wiesen eine durchschnittliche Berufserfahrung von 5,8 Jahren auf. Vgl. KIDA (1984), S. 334 bzw. TROTMAN/SNG (1989), S. 570.

<sup>6</sup> Vgl. CHURCH (1990), S. 89-91.

<sup>7</sup> So auch SCHREIBER (2000), S. 123 m.w.N.



Prüfungshandlungen zu sehen.<sup>1</sup> Im Gegensatz zu Going-Concern-Entscheidungen ist das mit einer einzelnen analytischen Prüfungshandlung verbundene Risiko weitaus geringer, da es bei Fehlern, die nicht mittels analytischer Prüfungshandlungen aufgedeckt werden konnten, weiterhin möglich ist, dass sie durch andere Prüfungshandlungen festgestellt werden können. Darüber hinaus könnten sich analytische Prüfungshandlungen auch deshalb als geeigneter erweisen, da sie im Rahmen von Jahresabschlussprüfungen regelmäßig zum Einsatz kommen. Going-Concern-Entscheidungen hingegen werden in der Regel nur bei Mandanten getroffen, bei denen Informationen vorliegen, die auf eine akute Gefährdung des Unternehmensfortbestands hinweisen. Abschließend bleibt noch festzuhalten, dass analytische Prüfungshandlungen im Gegensatz zu Going-Concern-Beurteilungen häufig von Prüfungsleitern durchgeführt werden, bei denen zu vermuten ist, dass sie aufgrund ihrer geringeren Anzahl an Berufsjahren tendenziell eher von Framing-Effekten betroffen sind als erfahrene Prüfer (etwa Manager oder Partner<sup>2</sup>).

Um Framing-Effekte bei analytischen Prüfungshandlungen zu untersuchen, ließen AYERS/KAPLAN 45 Prüfungsleiter mit einer durchschnittlichen Berufserfahrung von 3,3 Jahren eine analytische Prüfungshandlung im Rahmen der Prüfungsplanung durchführen.<sup>3</sup> Den Prüfern wurden ein Vorjahresvergleich auf der Grundlage von drei Kennzahlen (Bruttogewinnmarge, zweiter Liquiditätsgrad, dritter Liquiditätsgrad) sowie weitere Hintergrundinformationen eines hypothetischen Mandanten präsentiert. Die Analyse der Kennzahlen ergab in allen drei Fällen eine wesentliche Abweichung zwischen dem Wert des abgelaufenen Geschäftsjahrs und dem Wert des Vorjahrs.<sup>4</sup> Die Hintergrundinformationen umfassten eine Liste mit 18 fallbezogenen Hinweisen. Neun der Hinweise deuteten darauf hin, dass die wesentliche Abweichung auf einen Fehler im Rechnungswesen des Mandanten zurückzuführen sein könnte (sog. *misstatement*<sup>5</sup>). Aus den anderen neun Hinweisen konnte geschlossen werden, dass die Abweichung nicht durch einen Fehler im Rechnungswesen, sondern durch andere Faktoren verursacht wurde (sog. *non-misstatement*<sup>6</sup>).<sup>7</sup> Den teilnehmenden Prüfern wurde ausdrücklich mitgeteilt, dass

---

<sup>1</sup> Vgl. im Folgenden AYERS/KAPLAN (1993), S. 118.

<sup>2</sup> Der Partner steht auf der höchsten Stufe der Mitarbeiterhierarchie einer Prüfungsgesellschaft und ist gegenüber den Managern und Prüfungsleitern weisungsberechtigt. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 552.

<sup>3</sup> Vgl. AYERS/KAPLAN (1993), S. 118 i.V.m. 122.

<sup>4</sup> Vgl. AYERS/KAPLAN (1993), S. 119.

<sup>5</sup> Ein Beispiel für einen *misstatement* wäre etwa, dass langjährige Mitarbeiter der Debitorenabteilung während des abgelaufenen Geschäftsjahrs durch unerfahrenes Personal ersetzt wurden.

<sup>6</sup> Ein Beispiel für einen *non-misstatement* wäre etwa, dass der Mandant im abgelaufenen Geschäftsjahr einen neuen Großkunden akquirieren konnte.

<sup>7</sup> Vgl. AYERS/KAPLAN (1993), S. 121.

die wesentliche Abweichung sowohl durch ein *misstatement* als auch durch ein *non-misstatement* verursacht sein könnte.<sup>1</sup>

Bevor die Probanden die Liste mit den Hinweisen erhielten, wurde eine Gruppe der Prüfer aufgefordert, die Wahrscheinlichkeit zu schätzen, dass die wesentlichen Abweichungen in den Kennzahlen auf ein *misstatement* zurückzuführen ist. AYERS/KAPLAN gingen davon aus, dass allein die verwendete Wortwahl ein *misstatement-frame* induziert.<sup>2</sup> Anschließend sollten sie diejenigen Hinweise aus der Liste kennzeichnen, die ihrer Ansicht nach besonders in Betracht zu ziehen sind, um die festgestellte Abweichung in den Daten zu erklären. Die andere Gruppe sollte hingegen eine Wahrscheinlichkeitsbeurteilung für das Vorliegen eines *non-misstatements* vornehmen und daraufhin diejenigen Hinweise der Liste markieren, die besonders relevant für die Erklärung der Abweichung sind. Die Autoren vermuteten aufgrund der gewählten Begrifflichkeiten die Verursachung eines *non-misstatement-frames*.<sup>3</sup>

Anders als bei den Studien zu Going-Concern-Beurteilungen konnte festgestellt werden, dass die teilnehmenden Prüfungsleiter bei der Durchführung der analytischen Prüfungshandlung von Framing-Effekten betroffen waren. Im Einzelnen zeigten AYERS/KAPLAN, dass die Prüfer der *misstatement-frame*-Gruppe eine signifikant höhere Anzahl an Hinweisen auflistete, die auf das Vorliegen eines Fehlers deuten, und die Prüfer der *non-misstatement-frame*-Gruppe signifikant mehr Hinweise, die auf das Vorliegen eines externen Faktors weisen.<sup>4</sup> Die Autoren schlussfolgerten daraus, dass die Effizienz und Effektivität von analytischen Prüfungshandlungen möglicherweise durch die innerhalb eines Prüfungsstandards bzw. der Prüfungshandbücher gewählte Wortwahl beeinträchtigt werden kann. So könnte ein *non-misstatement-frame* dazu führen, dass ein Fehler erst durch spätere Einzelfallprüfungen aufgedeckt wird. Dieses könnte mit nicht unerheblichen Effizienzverlusten verbunden sein, da Einzelfallprüfungen mehr Zeit in Anspruch nehmen als analytische Prüfungshandlungen und dadurch höhere Kosten verursachen. Weiterhin könnte ein *non-misstatement-frame* im Rahmen der Prüfungsplanung dazu führen, dass der Umfang weiterer Prüfungshandlungen im betroffenen Prüffeld derart

---

<sup>1</sup> Vgl. AYERS/KAPLAN (1993), S. 119.

<sup>2</sup> Vgl. AYERS/KAPLAN (1993), S. 120.

<sup>3</sup> Vgl. AYERS/KAPLAN (1993), S. 120.

<sup>4</sup> Vgl. AYERS/KAPLAN (1993), S. 124.

eingeschränkt wird, dass der Fehler nicht mehr aufgedeckt werden kann, was die Effektivität der gesamten Jahresabschlussprüfung erheblich bedrohen könnte.<sup>1</sup>

Betrachtet man diese potenziellen Gefahren, so empfiehlt es sich, Prüfungshinweise, etwa in berufsrechtlichen Prüfungsnormen oder in Handbüchern der Prüfungsgesellschaften, mit einem *misstatement-frame* zu versehen.<sup>2</sup> Dieses verringert die potenzielle Gefahr, dass kognitive Verzerrungen zur Nicht-Aufdeckung von Fehlern führen.<sup>3</sup> Bezogen auf die konkreten, für die Durchführung von analytischen Prüfungshandlungen geltenden berufsrechtlichen Regelungen, kann festgestellt werden, dass weder IDW PS 312 noch ISA 520 auf Begrifflichkeiten zurückgreifen, die einen *non-misstatement-frame* induzieren könnten. Dennoch kann nicht ausgeschlossen werden, dass Prüfer bedingt durch Prüfungshinweise in den Handbüchern oder durch schulungsbedingte Vorgehensweisen einem *non-misstatement-frame* bei der Durchführung von analytischen Prüfungshandlungen unterliegen. Qualitätsmindernde Vorgehensweisen wären somit von den einzelnen Prüfungsgesellschaften abhängig.

#### **3.5.2.2.2 Anker-Effekt (Anchoring)**

##### **3.5.2.2.2.1 Die Berücksichtigung ungeprüfter Buchwerte im Rahmen der Erwartungsbildung**

Wie bereits unter Kap. 3.2.2.2.2 ausführlich dargestellt, wenden Individuen sog. *kognitive Daumenregeln* an, um die ihnen gegenüberstehende Komplexität reduzieren und somit handhaben zu können. Da auch die Jahresabschlussprüfung durch einen hohen Grad an Komplexität gekennzeichnet ist, dürften derartige Vereinfachungsregeln von großer Relevanz für das Verständnis der prüferischen Urteilsbildung sein.

Im Rahmen des Aufbaus einer Problemrepräsentation bei analytischen Prüfungshandlungen kommt insbesondere die Verwendung der Anker-Heuristik in Betracht.<sup>4</sup> Da die Prüfer im Rahmen der Jahresabschlussprüfung typischerweise bereits sehr früh über die ungeprüften Buchwerte des Mandanten verfügen, könnten sie diesen bei der Bildung eines Erwartungswerts als sog. *Anker* verwenden. In diesem Fall würde der entwickelte Erwartungswert stark von dem ungeprüften Buchwert beeinflusst werden. Nimmt man

---

<sup>1</sup> Vgl. AYERS/KAPLAN (1993), S. 127.

<sup>2</sup> Ein derartiger Vorschlag wurde bereits von KINNEY/HAYNES (1990), S. 95, 100 geäußert.

<sup>3</sup> So auch AYERS/KAPLAN (1993), S. 127.

<sup>4</sup> Vgl. zur Erläuterung der Anker-Heuristik ausführlich Kap. 3.2.2.2.2.

an, dass der ungeprüfte Buchwert des Mandanten jedoch einen wesentlichen Fehler enthält, so besteht zwangsläufig die Gefahr, dass der anschließende Soll-Ist-Vergleich keine wesentliche Abweichung signalisiert und somit das entsprechende Prüffeld irrtümlicherweise als korrekt oder zumindest als ein Prüffeld mit geringem Fehlerrisiko akzeptiert werden könnte.<sup>1</sup> Ein derartiger Einfluss des ungeprüften Buchwerts auf die prüferische Urteilsbildung würde demnach erhebliche negative Auswirkungen auf die Effektivität und Effizienz von Jahresabschlussprüfungen mit sich bringen.<sup>2</sup> Schließlich kann auch in Frage gestellt werden, ob eine solche prüferische Vorgehensweise mit den Berufsgrundsätzen des Prüfers vereinbar ist. Nach § 11 Abs. 1 der Berufssatzung hat ein Prüfer sein Handeln selbst zu bestimmen und seine Urteile selbst zu bilden. Da bei der Verwendung der Anker-Heuristik allerdings ein kausaler Zusammenhang zwischen dem entwickelten Erwartungswert des Prüfers und dem ungeprüften Buchwert des Mandanten nicht von der Hand zu weisen ist und folglich die Problemrepräsentation, welche Bestandteil der prüferischen Urteilsbildung ist, entscheidend vom Mandanten beeinflusst wird, bestehen erhebliche Zweifel bezüglich der Einhaltung der notwendigen Eigenverantwortlichkeit des Prüfers.<sup>3</sup>

Eine der ersten Studien, die sich mit der Frage beschäftigte, ob Prüfer bei der Durchführung von analytischen Prüfungshandlungen tatsächlich von den ungeprüften Werten des Mandanten beeinflusst werden, führten KINNEY/UECKER durch. Den insgesamt 179 teilnehmenden Prüfungsleitern<sup>4</sup> wurden hierzu die Umsatzrendite der beiden Vorjahre sowie des aktuellen Geschäftsjahrs zur Verfügung gestellt.<sup>1</sup> Dabei ist hervorzuheben, dass die beiden Vorjahreswerte auf Daten geprüfter Jahresabschlüsse basieren und der aktuelle Wert hingegen dem zu prüfenden Jahresabschluss entstammt. Die Aufgabe der Prü-

---

<sup>1</sup> Das Risiko der irrtümlichen Annahme wird in der Literatur auch als  $\beta$ -Fehler bezeichnet, dessen Beschränkung auf das maximal zulässige Niveau das eigentliche Hauptziel der Jahresabschlussprüfung darstellt. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 133-134.

<sup>2</sup> Dieser Ansicht sind auch weite Teile der Literatur. Vgl. etwa SCHREIBER (2000), S. 164-165; HEINTZ/WHITE/BEDARD (1999), S. 136; KENNEDY (1995), S. 250; WILD/BIGGS (1990), S. 228. Anderer Ansicht sind hingegen SHIELDS/SOLOMON/WALLER, die davon ausgehen, dass zwischen dem ungeprüften und dem tatsächlichen Buchwert eine starke Korrelation besteht. Demnach würde die Verwendung des Ersteren zur Schätzung des Letzteren die Effektivität und Effizienz des Prüfungsprozesses steigern. Vgl. SHIELDS/SOLOMON/WALLER (1988), S. 2 und 4. Nicht zuletzt aufgrund des fehlenden empirischen Nachweises ihrer Hypothese, wurde ihre Auffassung massiv kritisiert und in der weiteren wissenschaftlichen Diskussion nicht weiter berücksichtigt. Vgl. etwa WILD/BIGGS (1990), S. 239.

<sup>3</sup> Vgl. auch BIGGS/WILD (1985), S. 609; WILD/BIGGS (1990), S. 239 sowie HEINTZ/WHITE/BEDARD (1999), S. 136.

<sup>4</sup> Obwohl keine weiteren Angaben zu den genauen Berufsjahren der teilnehmenden Prüfer getroffen wurden, kann man davon ausgehen, dass die an empirischen Studien teilnehmenden Prüfungsleiter durchschnittlich über drei Jahre Berufserfahrung verfügen.

fer bestand darin, eine Bandbreite von möglichen Werten für das aktuelle Geschäftsjahr anzugeben, bei deren Überschreitung sie weitere Untersuchungen der zugrunde liegenden Jahresabschlusspositionen in Erwägung ziehen würden (sog. investigation boundaries). Um einen möglichen Einfluss des ungeprüften Werts feststellen zu können, erhielt die eine Gruppe der Probanden einen hohen ungeprüften Wert und die andere Gruppe der Prüfer einen niedrigeren ungeprüften Wert. Die beiden Vorjahreswerte hingegen waren in beiden Gruppen identisch.<sup>2</sup>

Ein Vergleich der beiden Prüfergruppen zeigte, dass die angegebenen investigation boundaries signifikant voneinander abwichen.<sup>3</sup> Die Prüfer, die einen niedrigen ungeprüften Wert erhielten, setzten niedrigere investigation boundaries als Prüfer, denen ein hoher ungeprüfter Wert zur Verfügung gestellt wurde. Somit ließen die Ergebnisse vermuten, dass die angegebenen Bandbreiten in Richtung des ungeprüften Werts beeinflusst wurden, was auf das Vorliegen des Anker-Effekts deutet. Allerdings stellten KINNEY/UECKER ebenfalls fest, dass die Beeinflussung durch den hohen ungeprüften Wert weniger stark ausgeprägt war als durch den niedrigeren ungeprüften Wert. Die Autoren führten diese Beobachtung darauf zurück, dass in Abhängigkeit vom ungeprüften Wert der Trend in den Vorjahreswerten entweder bestätigt oder durchbrochen wurde. Da die beiden Vorjahreswerte durch einen fallenden Trend gekennzeichnet waren, stellte der niedrige ungeprüfte Wert eine Fortführung des negativen Trends dar, wohingegen der hohe Wert eine Trendwende verursachte. Letztgenanntes könnte eine erhöhte Aufmerksamkeit der betroffenen Prüfer bewirkt und dem Anker-Effekt zum Teil entgegengewirkt haben.

Um den Zusammenhang zwischen dem Abwärtstrend in den Vorjahren und dem ungeprüften Wert des aktuellen Jahrs genauer zu untersuchen, führten BIGGS/WILD eine Erweiterungsstudie zu KINNEY/UECKER durch.<sup>4</sup> BIGGS/WILD waren der Ansicht, dass das Untersuchungsdesign von KINNEY/UECKER lediglich dazu geeignet war, das Vorhandensein des Anker-Effekts nachzuweisen. Eine Beurteilung des Ausmaßes in Abhän-

---

<sup>1</sup> Bei der Umsatzrendite handelt es sich um eine Kennzahl, die die Umsatzerlöse in Relation zu den Umsatzkosten setzt. Beide Positionen können dem Jahresabschluss entnommen werden.

<sup>2</sup> Vgl. KINNEY/UECKER (1982), S. 57-58, 61.

<sup>3</sup> Vgl. hier und im Folgenden KINNEY/UECKER (1982), S. 61-62.

<sup>4</sup> Das heißt, dass die von BIGGS/WILD verwendete Prüfungsaufgabe identisch zu der Studie von KINNEY/UECKER ist, das Untersuchungsdesign hingegen an den entscheidenden Stellen modifiziert bzw. erweitert wurde, um Schwächen des ursprünglichen Designs zu eliminieren oder um weitere Aussagen treffen zu können. Insgesamt nahmen an der Studie von BIGGS/WILD 121 Prüfer teil, deren Anzahl an Praxisjahren von weniger als zwei bis über 24 Jahre variierte. Interessanterweise konnten die Autoren keine Unterschiede in der Existenz und dem Ausmaß von Anker-Effekten in Abhängigkeit von der Berufserfahrung feststellen. Vgl. BIGGS/WILD (1985), S. 613.

gigkeit von der Art des zur Verfügung gestellten dritten Werts (hoher ungeprüfter Wert/Trendwende vs. niedriger ungeprüfter Wert/Trendfortführung) konnte die Studie jedoch nicht erbringen. Demzufolge war es bei den Prüfern mit dem niedrigen Wert schwierig zu differenzieren, ob die angegebenen investigation boundaries durch den ungeprüften Wert beeinflusst wurden oder durch den Trend in den Vorjahreswerten.<sup>1</sup> Zur Beantwortung dieser Frage ließen BIGGS/WILD neben der Gruppe mit dem hohen und der Gruppe mit dem niedrigen ungeprüften Wert (sog. experimental groups) eine weitere Gruppe teilnehmen, die sich aus Prüfern zusammensetzte, welche lediglich die beiden Vorjahreswerte, nicht jedoch den ungeprüften Wert des aktuellen Geschäftsjahrs erhielten (sog. control group).<sup>2</sup> Da diese demnach nicht durch einen ungeprüften Wert beeinflusst werden konnten, gingen die Autoren davon aus, dass ein Vergleich der experimental groups mit der control group zusätzliche Informationen über das Ausmaß des Einflusses des ungeprüften Werts zur Verfügung stellt.<sup>3</sup>

Die Ergebnisse von BIGGS/WILD konnten die Ergebnisse von KINNEY/UECKER zunächst einmal bestätigen. Zwischen den investigation boundaries der Prüfer, denen ein hoher ungeprüfter Wert bereitgestellt wurde (im Folgenden als experimental group 1 bezeichnet), und den Prüfern denen ein niedriger ungeprüfter Wert zur Verfügung gestellt wurde (im Folgenden als experimental group 2 bezeichnet) konnten signifikante Unterschiede festgestellt werden, wobei die Angaben in Richtung des ungeprüften Werts beeinflusst wurden. Folglich sind auch BIGGS/WILD der Ansicht, dass im Rahmen von analytischen Prüfungshandlungen eine potenzielle Gefahr von der Kenntnis des ungeprüften Werts ausgeht.<sup>4</sup>

In einer weiteren Analyse verglichen BIGGS/WILD die investigation boundaries der experimental groups mit denen der control group. Hierbei konnte festgestellt werden, dass die Unterschiede zwischen der experimental group 1 und der control group signifikant größer waren als die Unterschiede zwischen der experimental group 2 und der control group. Hieraus lassen sich zwei wichtige Erkenntnisse ableiten. Erstens trafen die Prüfer

---

<sup>1</sup> Vgl. BIGGS/WILD (1985), S. 612.

<sup>2</sup> Eine zusätzliche Erweiterung zu KINNEY/UECKER bestand darin, dass BIGGS/WILD die Anzahl der gegebenen Vorjahreswerte variierten. Neben den Prüfergruppen, denen die Werten aus zwei Vorjahren zur Verfügung gestellt wurden, bildeten sie weitere Prüfergruppen, denen zur Bearbeitung der gleichen Aufgabe insgesamt fünf Vorjahreswerte bereitgestellt wurden. Hierdurch versuchten die Autoren herauszufinden, ob die Existenz von Anker-Effekten von der Anzahl der zur Verfügung gestellten Vorjahreswerte abhängt. Da jedoch kaum Unterschiede zu den Gruppen mit zwei Vorjahreswerten festgestellt werden konnten, wird dieser Aspekt hier nicht weiter erörtert. Vgl. BIGGS/WILD (1985), S. 615-616, 621.

<sup>3</sup> Vgl. BIGGS/WILD (1985), S. 612-613.

<sup>4</sup> Vgl. BIGGS/WILD (1985), S. 616-617.

mit dem niedrigen ungeprüften Wert ähnliche Urteile wie die Prüfer ohne ungeprüften Wert. BIGGS/WILD führten diese Beobachtung darauf zurück, dass die Entscheidungen der Prüfer ohne ungeprüften Wert konsistent zum Abwärtstrend in den Vorjahresdaten waren.<sup>1</sup> Und zweitens wurden die Prüfer mit dem hohen Wert wesentlich stärker von dem ungeprüften Wert beeinflusst als die Prüfer mit dem niedrigen Wert.<sup>2</sup> Demnach sind die Schlussfolgerungen von KINNEY/UECKER ernsthaft zu bezweifeln. Denn wie die Ergebnisse von BIGGS/WILD zeigen, war es in der Gruppe mit dem niedrigen ungeprüften Buchwert nicht dem Anker-Effekt zuzurechnen, dass die investigation boundaries in die Richtung des ungeprüften Werts gelenkt wurden, sondern dem Abwärtstrend in den Vorjahresdaten. Quasi im Umkehrschluss ist zu revidieren, dass die Trendwende bei der Gruppe mit dem hohen ungeprüften Wert eine erhöhte Aufmerksamkeit auslöste und somit dem Anker-Effekt entgegenwirkte. Vielmehr deuten die Ergebnisse von BIGGS/WILD darauf hin, dass der mit einer Trendwende verbundene hohe Wert zu einer stärkeren Ausrichtung der Erwartungen an den ungeprüften Wert führt, als der mit einer Trendfortführung verbundene niedrigere Wert.

Eine weitere Studie, die als Erweiterung zu KINNEY/UECKER und BIGGS/WILD zu sehen ist, führten HEINTZ/WHITE durch.<sup>3</sup> Sie weisen darauf hin, dass den bisherigen Studien lediglich ein fallender Trend in den Vorjahreswerten zugrunde gelegt wurde. Daher ließen die beiden vorangegangenen Studien folgende drei Fragen unbeantwortet:

1. Können die Erkenntnisse von KINNEY/UECKER und BIGGS/WILD auch auf einen steigenden Vorjahrestrend übertragen werden oder gelten sie ausschließlich für sinkende Vorjahreswerte?
2. Übt ein hoher ungeprüfter Wert, der zu einer Trendwende (nach oben) führt, einen stärkeren Einfluss auf den Prüfer aus als ein niedriger ungeprüfter Wert, der eine Trendwende (nach unten) einleitet?
3. Übt ein hoher ungeprüfter Wert, der den Trend (nach oben) fortführt, einen stärkeren Einfluss auf den Prüfer aus als ein niedriger Wert, der den Trend (nach unten) fortsetzt?

---

<sup>1</sup> Vgl. BIGGS/WILD (1985), S. 617.

<sup>2</sup> Vgl. BIGGS/WILD (1985), S. 617-618.

<sup>3</sup> Die Prüfungsaufgabe bestand wiederum darin, einen Erwartungswert sowie eine obere und eine untere Grenze für die Umsatzrendite des aktuellen Geschäftsjahrs anzugeben. Insgesamt nahmen an dieser Studie 140 Prüfer teil, die über eine Berufserfahrung zwischen 18 Monaten und 23 Jahren verfügten. Wie BIGGS/WILD konnten auch HEINTZ/WHITE keine signifikanten Unterschiede in der Existenz und

Zur Beantwortung dieser Fragen integrierten HEINTZ/WHITE neben dem fallenden Trend zusätzlich einen steigenden Trend in das Untersuchungsdesign.<sup>1</sup> Insgesamt bildeten sie sechs Gruppen, wobei die Gruppen 1 bis 3 identisch zu BIGGS/WILD waren.<sup>2</sup> Gruppe 4 bis 6 erhielten hingegen Vorjahresdaten, die im Gegensatz zu BIGGS/WILD durch einen steigenden Trend gekennzeichnet waren. Gruppe 4 wurde zusätzlich ein (hoher) ungeprüfter Wert zur Verfügung gestellt, der den steigenden Trend der Vorjahre fortsetzte (experimental group 3), Gruppe 5 hingegen ein (niedriger) ungeprüfter Wert, der eine Trendwende darstellte (experimental group 4). Gruppe 6 erhielt keinen ungeprüften Wert (control group 2).<sup>3</sup>

Vorweg ist zunächst einmal festzuhalten, dass HEINTZ/WHITE die Erkenntnisse der beiden vorangegangenen Studien (KINNEY/UECKER und BIGGS/WILD) bestätigen konnten. Die Ergebnisse wiesen darauf hin, dass ein signifikanter Einfluss von den ungeprüften Werten auf die Erwartungen der Prüfer ausgeübt wurde, und zwar unabhängig davon, ob die Vorjahresdaten durch einen steigenden oder fallenden Trend geprägt waren oder der ungeprüfte Wert eine Trendfortsetzung oder -wende darstellte.<sup>4</sup> Darüber hinaus zeigten die Ergebnisse auch, dass die teilnehmenden Prüfer signifikant stärker von dem ungeprüften Wert beeinflusst wurden, wenn der Wert eine Trendwende einleitete. Insofern konnten die Autoren auch die über KINNEY/UECKER hinausgehenden Ergebnisse von BIGGS/WILD bestätigen. Hervorzuheben ist allerdings, dass BIGGS/WILD dieses ausschließlich für einen fallenden Vorjahrestrend feststellen konnten. HEINTZ/WHITE erweitern deren Erkenntnisse dahingehend, dass ihre Ergebnisse sowohl in einem sinken-

---

dem Ausmaß des Framing-Effekts in Abhängigkeit der Berufserfahrung feststellen. Vgl. HEINTZ/WHITE (1989), S. 24-26.

<sup>1</sup> Vgl. HEINTZ/WHITE (1989), S. 22-23. Neben der zusätzlichen Berücksichtigung eines steigenden Trends in den Vorjahresdaten, nahmen HEINTZ/WHITE eine weitere Modifikation zu WILD/BIGGS vor. Da sie der Auffassung waren, dass die teilnehmenden Prüfer bei KINNEY/UECKER und BIGGS/WILD den ungeprüften Wert möglicherweise gar nicht als ungeprüft wahrgenommen haben (zu den einzelnen Gründen siehe HEINTZ/WHITE (1989), S. 23-24 sowie 26-27), gaben sie den Probanden nicht den Gesamtjahreswert der aktuellen Periode, sondern lediglich einen neun-Monatswert. Die Autoren hielten es für möglich, dass die Probanden bei BIGGS/WILD sowie KINNEY/UECKER deshalb von dem ungeprüften Wert beeinflusst wurden, weil sie diesen als einen geprüften Wert wahrgenommen haben. Im Umkehrschluss vermuten HEINTZ/WHITE, dass die Prüfer in ihrer Studie gar nicht von dem ungeprüften Wert beeinflusst werden. Diese Annahme konnte jedoch nicht bestätigt werden. Vgl. HEINTZ/WHITE (1989), S. 31.

<sup>2</sup> Alle drei Gruppen erhielten die Umsatzrenditen der beiden letzten Geschäftsjahre, wobei sie durch einen fallenden Trend gekennzeichnet waren. Gruppe 1 bekam zusätzlich einen (niedrigen) ungeprüften Wert des aktuellen Geschäftsjahrs, der den sinkenden Trend fortführte (experimental group 1), Gruppe 2 hingegen einen (hohen) ungeprüften Wert, der eine Trendwende darstellte (experimental group 2). Der Gruppe 3 wurde kein ungeprüfter Wert bereitgestellt (control group 1). Vgl. HEINTZ/WHITE (1989), S. 24.

<sup>3</sup> Vgl. HEINTZ/WHITE (1989), S. 24.

<sup>4</sup> Vgl. HEINTZ/WHITE (1989), S. 31.



den als auch in einem steigenden Verlauf beobachtet wurden.<sup>1</sup> Somit konnte die zuvor aufgeworfene Frage 1 an dieser Stelle beantwortet werden.

Zur Klärung der Frage 2 stellten die Autoren die Ergebnisse der experimental group 2 (hoher Wert/Trendwende nach oben) und der experimental group 4 (niedriger Wert/Trendwende nach unten) gegenüber. Die Ergebnisse zeigten, dass der niedrigere ungeprüfte Wert, der eine Trendwende nach unten einleitete, einen leicht stärkeren Einfluss auf den Prüfer hatte als ein höherer ungeprüfter Wert, der eine Trendwende nach oben bewirkte.<sup>2</sup> Ebenso verhielt es sich bei der Beantwortung der Frage 3. Hierzu bedurfte es eines Vergleichs zwischen der experimental group 1 (niedriger Wert/Trendfortführung) und der experimental group 3 (hoher Wert/Trendfortführung). Auch hier konnte ein leicht stärkerer Einfluss des niedrigeren ungeprüften Werts festgestellt werden.<sup>3</sup>

Zusammenfassend bleibt festzuhalten, dass aus den Studien von KINNEY/UECKER, WILD/BIGGS und HEINTZ/WHITE drei wesentliche Erkenntnisse zum Einfluss ungeprüfter Buchwerte auf die prüferische Urteilsbildung gewonnen werden konnten, die im Folgenden kurz wiederholt und kritisch diskutiert werden sollen:

- 1) Die Kenntnis über den ungeprüften Buchwert beeinflusst die Erwartungsbildung des Prüfers in die Richtung des ungeprüften Werts. Dieser sog. Anker-Effekt konnte in allen drei Studien festgestellt werden, so dass die Erkenntnis als empirisch gesichert betrachtet werden kann.<sup>4</sup>
- 2) Der Prüfer wird leicht stärker von einem ungeprüften Wert beeinflusst, wenn es sich dabei um einen (niedrigen) Wert handelt, der einen fallenden Trend fortführt, als wenn es sich dabei um einen (hohen) Wert handelt, der einen steigenden Trend fortsetzt.
- 3) Der Prüfer wird wesentlich stärker von einem ungeprüften Wert beeinflusst, wenn der ungeprüfte Wert eine Trendwende darstellt, unabhängig davon, ob ein sinkender Trend nach oben oder ein steigender Trend nach unten durchbrochen wird.

---

<sup>1</sup> Vgl. HEINTZ/WHITE (1989), S. 32.

<sup>2</sup> Vgl. HEINTZ/WHITE (1989), S. 32.

<sup>3</sup> Vgl. HEINTZ/WHITE (1989), S. 32.

<sup>4</sup> HEINTZ/WHITE/BEDARD (1999), S. 141 konnten diese Feststellung in einer erneuten Erweiterungsstudie zu HEINTZ/WHITE (1989) wiederholt bestätigen. KENNEDY (1995), S. 263 führte eine vergleichbare Studie durch und konnte den Einfluss von ungeprüften Buchwerten in einem modifizierten Untersuchungsdesign ebenfalls bestätigen. Da das Hauptanliegen der beiden Studien allerdings auf das Ent-

Die erste Erkenntnis impliziert erhebliche negative Auswirkungen auf die Effektivität und Effizienz von analytischen Prüfungshandlungen und somit auf den gesamten Prüfungsprozess. Sollte ein ungeprüfter Wert einen wesentlichen Fehler enthalten, so besteht einerseits die Gefahr, dass Anker-Effekte die Erwartungsbildung in Richtung dieses Werts lenken und die Divergenz zum Ist-Wert nicht mehr als wesentlich eingestuft wird. Eine unwesentliche Soll-Ist-Abweichung könnte den Prüfer dazu verleiten, die Prüfungshandlungen für das entsprechende Prüffeld auszusetzen, auf ein Minimum zu beschränken oder zumindest nicht auszudehnen. Dieses könnte dazu führen, dass Fehler unentdeckt bleiben, was das Risiko der Annahme eines mit wesentlichen Fehlern behafteten Prüffelds (beta-Risiko) erheblich erhöht. Zwar könnten diese Fehler auch noch zu einem späteren Zeitpunkt durch andere Prüfungshandlungen, etwa Einzelfallprüfungen, aufgedeckt werden. Da diese im Vergleich zu analytischen Prüfungshandlungen jedoch mit erheblich mehr Zeitaufwand und Kosten verbunden sind, läge hier immer noch eine Beeinträchtigung der Wirtschaftlichkeit des Prüfungsauftrags vor. Zwar könnte dem entgegengehalten werden, dass Anker-Effekte bei einem Prüffeld ohne wesentliche Fehler das Risiko einer irrtümlichen Ablehnung (alpha-Risiko) erhöhen. Jedoch könnten dieser Argumentation drei entscheidende Gegenargumente entgegengebracht werden. Erstens ist das Hauptziel einer Jahresabschlussprüfung in der Beschränkung des beta- bzw. Prüfungsrisikos auf das maximal zulässige Niveau zu sehen.<sup>1</sup> Somit sollten die Prüfungshandlungen auf die Entdeckung wesentlicher Fehler ausgerichtet werden und nicht auf die Vermeidung „falscher Signale“.<sup>2</sup> Zweitens würde das Anpassen der Erwartungshaltung des Prüfers an den ungeprüften Buchwert gleichzeitig bedeuten, dass man auf die Richtigkeit des Buchwerts vertraut, ohne zuvor eine Prüfungshandlung vorgenommen zu haben. Dieses stellt die Notwendigkeit der Abschlussprüfung erheblich in Frage.<sup>1</sup> Und drittens kann die Beeinflussung der prüferischen Urteilsbildung durch eine Aussage des Mandanten (hierzu gehört selbstverständlich auch der ungeprüfte Buchwert) sicherlich als ein Verstoß gegen den Grundsatz der Eigenverantwortlichkeit gewertet werden.

Die zweite Erkenntnis impliziert, dass sich Prüfer leicht stärker durch Unterbewertung beeinflussen lassen als durch Überwertungen. Diese Feststellung ist (zumindest bezogen auf die Umsatzrendite) weniger bedrohlich für die Jahresabschlussprüfung. Nach allge-

---

gegenwirken von Framing-Effekten (sog. debiasing) liegt, erfolgt eine ausführliche Darstellung unter Kap. 3.5.2.2.2.2.

<sup>1</sup> Vgl. WOLZ (2003), S. 8.

<sup>2</sup> Ähnlich argumentieren BIGGS/WILD (1985), S. 611.

meiner Auffassung sollten Prüfer verstärkt Fehler identifizieren, die eine Überbewertung verursachen.<sup>2</sup> Diese Schlussfolgerung kann ebenfalls auf Positionen der Aktivseite übertragen werden, da auch hier Überbewertungen ein höheres Prüfungsrisiko mit sich bringen als Unterbewertungen. Problematisch ist allerdings, wenn diese Erkenntnis auch auf analytische Prüfungshandlungen übertragen werden können, die auf der Passivseite oder im Bereich der Aufwendungen durchgeführt werden. Hier geht für die Jahresabschlussprüfung ein höheres Risiko von Unter- als von Überbewertungen aus.

Die dritte Erkenntnis impliziert, dass sich Prüfer genau dann besonders stark von den ungeprüften Buchwerten des Mandanten beeinflussen lassen, wenn es nicht der Fall sein sollte. Denn Trendumbrüche sind Zeichen für Änderungen im Geschäftsverlauf, bei denen der Prüfer besonders gründlich und vorsichtig vorgehen sollte.<sup>3</sup> Demnach liegt auch hier eine ernstzunehmende Gefahr für die Effektivität und Effizienz des Prüfungsprozesses vor.

Zusammenfassend ist herauszustellen, dass die Kenntnis über die ungeprüften Buchwerte des Mandanten aus zwei Überlegungen heraus kritisch zu beurteilen ist. Erstens könnte ihre Verwendung die Effektivität und Effizienz von analytischen Prüfungshandlungen erheblich beeinträchtigen, und zweitens ist die Einhaltung des Grundsatzes der Eigenverantwortlichkeit mit einer potenziellen Beeinflussung durch den ungeprüften Buchwert des Mandanten schwierig zu vereinbaren. Unter diesen Umständen ist die Vorgehensweise in der Prüfungspraxis ernsthaft in Frage zu stellen. Die Suche nach einem Weg zur Vermeidung derartiger Missstände scheint dringend notwendig. Konsequenterweise wäre zu empfehlen, dass Prüfer ihre Erwartungsbildung ohne Kenntnis des Buchwerts durchführen sollten. Derartige Handlungsempfehlungen könnten etwa in Prüfungsstandards bzw. -handbücher, Vorlesungen oder Lehrbücher sowie im Rahmen von Schulungen der Prüfungsgesellschaften aufgenommen werden. Allerdings ist fraglich, ob eine derartige Vorgehensweise praktisch tatsächlich umsetzbar ist, denn die Übernahme der Jahresabschlussdaten des Mandanten in die eigene Prüfungssoftware ist eine der ersten Handlungen des Prüfungsleiters. Sie dienen als Grundlage der Ermittlung von Materiality-Werten und der ersten Risikoeinschätzung im Rahmen der Prüfungsplanung.

---

<sup>1</sup> Vgl. WILD/BIGGS (1990), S. 239.

<sup>2</sup> Vgl. HEINTZ/WHITE (1989), S. 36.

<sup>3</sup> Vgl. HEINTZ/WHITE (1989), S. 36.

Geht man davon aus, dass die Durchführung analytischer Prüfungshandlungen ohne die Kenntnis über den ausgewiesenen Buchwert in der Prüfungspraxis kaum umzusetzen ist, müssen also Lösungsansätze gefunden werden, die nicht die Ursache, also den ungeprüften Buchwert, beseitigen, sondern die Auswirkungen, also die Beeinflussung des Prüfers durch den ungeprüften Buchwert, vermeiden. Studien zur Entdeckung von Möglichkeiten der „Immunisierung“ des Prüfers werden als *Debiasing*-Studien bezeichnet, auf die im folgenden Abschnitt genauer eingegangen werden soll.

#### **3.5.2.2.2.2 Debiasing als Möglichkeit zur Vermeidung des Anker-Effekts**

Eine der ersten Studien, die sich mit dem *debiasing* von Anker-Effekten im Rahmen von analytischen Prüfungshandlungen beschäftigten, führte KENNEDY durch. Sie ordnet die unbewusste Ausrichtung des Prüfers an einen ungeprüften Buchwert dem sog. „curse of knowledge“ zu.<sup>1</sup> Der „curse of knowledge“ bezeichnet die Unfähigkeit des Prüfers, Information bzw. Wissen nicht in Betracht zu ziehen, die er bereits verarbeitet hat.<sup>2</sup> Der Grund hierfür ist darin zusehen, dass Individuen alle erlangten Informationen, und hierzu gehört auch ein ungeprüfter Buchwert, sachverhaltsbezogen speichern. Die für eine Erwartungsbildung erforderliche mentale Repräsentation des Sachverhalts greift anschließend auf alle erhaltenen Informationen, und somit auch auf den ungeprüften Buchwert, zurück. Da sich dieser mentale Prozess unbewusst vollzieht, ist sich der Prüfer dabei nicht im Klaren, dass sich der ungeprüfte Wert auf seine Urteilsbildung auswirkt.<sup>3</sup>

KENNEDY entwickelte einen *debiasing*-Bezugsrahmen und geht davon aus, dass Verzerrungen (*biases*) entweder *effort-related* oder *data-related* sind.<sup>1</sup> *Effort* besteht aus zwei Komponenten: Kapazität und Motivation. KENNEDY nimmt an, dass Prüfer über ausreichende geistige Kapazitäten verfügen und dass insbesondere fehlende Motivation für Verzerrungen verantwortlich ist. Fehlende Motivation äußert sich in mangelnder Aufmerksamkeit für die zu bearbeitende Aufgabe und ungenügenden mentalen Anstrengungen (*cognitive effort*). Ein angemessenes Mittel zur Erhöhung der Motivation ist die

---

<sup>1</sup> Der curse of knowledge konnte bereits in unterschiedlichen Prüfungsaufgaben beobachtet werden. Neben analytischen Prüfungshandlungen kommen insbesondere auch die Beurteilung des internen Kontrollsystems (vgl. etwa. REIMERS/BUTLER (1992), S. 185-194) oder Going-Concern-Entscheidungen (vgl. etwa CLARKSON/EMBY/WATT (2002), S. 7-20 sowie EMBY/GELARDI/LOWE (2002), S. 87-103) in Betracht.

<sup>2</sup> Vgl. KENNEDY (1995), S. 249.

<sup>3</sup> Vgl. KENNEDY (1995), S. 254.

Schaffung von Anreizen. Hierzu zählen etwa finanzielle Anreize oder der Aufbau von Verantwortungsdruck (*accountability*). Im Prüfungsprozess wird Verantwortungsdruck insbesondere durch den Review-Prozess aufgebaut.

Weiterhin können Verzerrungen aufgrund von unzureichenden Daten entstehen, also *data-related* sein. Hierbei kann zwischen internen und externen Daten differenziert werden. Während mit internen Daten das abrufbare Wissen des Individuums gemeint ist, beziehen sich externe Daten auf die Informationen und Signale aus der Umwelt. Verzerrungen, die aufgrund von unzureichenden internen Daten entstehen, können etwa durch Auffrischung des Wissens, Bereitstellung von Entscheidungshilfen, Fortbildung oder Prüfer austausch beseitigt werden. Verzerrungen aufgrund von unzureichenden externen Daten hingegen durch die Beschaffung zusätzlicher Informationen, die Ausarbeitung, Verdeutlichung und Umformulierung vorhandener Informationen oder die Eliminierung überflüssiger Informationen. Allerdings ist eine klare Trennung zwischen internen und externen Daten nicht immer durchführbar. Wie bereits oben gezeigt, entstammt etwa der ungeprüfte Buchwert der Umwelt und wird dann vom Prüfer mental verarbeitet, so dass es sich hierbei sowohl um eine interne als auch um eine externe Information handelt.

Um festzustellen, ob der Anker-Effekt *effort-related* oder *data-related* ist, untersucht KENNEDY die Auswirkung von drei potenziellen *debiasing*-Mechanismen:<sup>2</sup>

- Verantwortungsdruck (*accountability*);
- Erfahrung;
- Gegenbegründungen (*counterexplanation*).

„Accountability is an effort-inducing incentive and should mitigate effort-related biases“.<sup>3</sup> Erfahrung bezieht sich auf das vorhandene Wissen eines Prüfers und daher auf interne Daten. „Counterexplanations (...) involves explicitly considering evidence that would not support or lead one to expect the outcome that occurred. (...). Specifically, it weakens causal connections between evidence and the actual outcome. Alternative outcomes are made more salient with counterexplanation“.<sup>4</sup> *Counterexplanations* beziehen

---

<sup>1</sup> Vgl. im Folgenden KENNEDY (1995), S. 251-252 sowie KENNEDY (1993), S. 233.

<sup>2</sup> Vgl. KENNEDY (1995), S. 249.

<sup>3</sup> KENNEDY (1995), S. 250.

<sup>4</sup> KENNEDY (1995), S. 250.

sich somit auf externe Daten, die in erster Linie darauf ausgerichtet sind, bereits vorhandene Informationen auszuarbeiten und zu verdeutlichen.

Um die Möglichkeiten eines *debiasing* anhand der o.a. Mechanismen zu untersuchen, führte KENNEDY zwei Experimente durch, an denen insgesamt 229 Studenten und 322 Manager<sup>1</sup> teilnahmen. Das erste Experiment (im Folgenden als Experiment 1 bezeichnet) diente der Untersuchung des Einflusses von *accountability* sowie Erfahrung, das zweite Experiment (im Folgenden als Experiment 2 bezeichnet) hingegen der Untersuchung von *counterexplanations*. In beiden Experimenten erhielten die Probanden eine graphische Darstellung der Verkaufszahlen eines Produkts X (steigender Trend) bzw. Produkts Y (fallender Trend) für die vergangenen zwölf Perioden. Die Aufgabe der Probanden bestand darin, die Anzahl der verkauften Einheiten für die 13. Periode zu schätzen.<sup>2</sup>

Experiment 1 basierte auf einem  $2 \times 3 \times 2$  Untersuchungsdesign, mit den Merkmalen *accountability*, ungeprüfter Wert und Gruppenzugehörigkeit. Die beiden *accountability*-Ausprägungen waren mit oder ohne Verantwortungsdruck, wobei den Probanden mit *accountability* mitgeteilt wurde, dass sie sich für ihre Urteile zu einem späteren Zeitpunkt rechtfertigen müssen. In Bezug auf das zweite Merkmal (ungeprüfter Wert) wurde den Probanden entweder ein hoher, ein niedriger oder kein ungeprüfter Wert zur Verfügung gestellt. Die Gruppenzugehörigkeit differenziert sich in Studenten und Prüfer.<sup>3</sup>

Obwohl sich *accountability* und Erfahrung in diversen Studien als mögliche *debiasing*-Mechanismen herausgestellt haben,<sup>4</sup> konnte KENNEDY in ihrer Studie keine derartige Auswirkung feststellen. Zwischen Studenten und Managern konnten keine signifikanten Unterschiede in Bezug auf die Beeinflussung durch den zur Verfügung gestellten Wert identifiziert werden, so dass festgehalten werden kann, dass Anker-Effekte unabhängig von dem Grad der Erfahrung aufgetreten sind. Darüber hinaus konnte dem Anker-Effekt auch nicht durch eine gesteigerte Verantwortung entgegengewirkt werden, so dass auch hier gesagt werden kann, dass *accountability* nicht geeignet scheint, um Verzerrungen entgegenzuwirken.<sup>5</sup>

---

<sup>1</sup> Obwohl keine genaueren Angaben über die Berufsjahre getroffen wurden, ist davon auszugehen, dass Manager in der Regel mindestens 5-6 Jahre Praxiserfahrung aufweisen.

<sup>2</sup> Vgl. KENNEDY (1995), S. 260-262, 267.

<sup>3</sup> Vgl. KENNEDY (1995), S. 261.

<sup>4</sup> Vgl. SMITH/KIDA (1991), S. 472-489; KENNEDY (1993), S. 231-245; TAN (1995), S. 113-135.

<sup>5</sup> Vgl. KENNEDY (1995), S. 263-266.

Das anschließende Experiment 2 diente ausschließlich der Untersuchung der Auswirkungen von *counterexplanations*. Der Aufbau des Experiments war vergleichbar zum ersten Experiment, mit der Ausnahme, dass *counterexplanations* anstelle von *accountability* in das Untersuchungsdesign integriert wurden. Probanden in der *counterexplanation*-Gruppe (experimental group) wurden gebeten, Gegenargumente für das Eintreten des hohen bzw. niedrigen ungeprüften Werts zu dokumentieren, bevor sie eine Schätzung für die Verkaufszahlen der 13. Periode abgaben. Die andere Gruppe (control group) wurde ohne derartige Anweisungen zur Abgabe eines Schätzwerts instruiert.<sup>1</sup> Die Ergebnisse von KENNEDY zeigten, dass die Schätzungen der Probanden ohne *counterexplanations* (control group) im Vergleich zu den Schätzungen der Probanden mit *counterexplanation* (experimental group) wesentlich stärker in Richtung des ungeprüften Werts beeinflusst wurden. Demzufolge ist festzuhalten, dass *counterexplanation* eine potenzielle Eignung zur Mäßigung von Anker-Effekten zugesprochen werden kann.<sup>2</sup>

Eine weitere Studie, die dem *debiasing* von Anker-Effekten zuzuordnen ist, führten HEINTZ/WHITE/BEDARD durch. HEINTZ/WHITE äußerten bereits in einer Vorgängerstudie die Befürchtung, dass die teilnehmenden Prüfer immer dann von dem Anker-Effekt betroffen werden könnten, wenn der Aufgabenstellung keine Hinweise auf eine fehlende Zuverlässigkeit der Daten des Mandanten, etwa durch ein erhöhtes Prüfungsrisiko, entnommen werden konnte. Nach Einschätzung der Autoren könnten derartige Umstände dazu führen, dass die teilnehmenden Prüfer dem ungeprüften Buchwert einen gewissen Wahrheitsgehalt zuweisen und deshalb von ihnen beeinflusst werden.<sup>3</sup> Keine der Studien, in denen Anker-Effekte empirisch nachgewiesen werden konnten, enthielt Hinweise in der Aufgabenstellung, die auf eine mangelnde Zuverlässigkeit der ungeprüften Buchwerte hindeuteten.<sup>4</sup> HEINTZ/WHITE/BEDARD sind jedoch der Auffassung, dass Informationen, die auf eine mangelnde Zuverlässigkeit der Finanzbuchführung hinweisen, das Vertrauen in die ungeprüften Buchwerte und somit das Auftreten von Framing-Effekten entscheidend mindern könnten.<sup>5</sup> Um einen derartigen *debiasing*-

---

<sup>1</sup> Vgl. KENNEDY (1995), S. 267.

<sup>2</sup> Vgl. KENNEDY (1995), S. 267-270.

<sup>3</sup> Vgl. HEINTZ/WHITE (1989), S. 22-23, 37. Die Autoren führten im Anschluss an ihre Studie Interviews mit den teilnehmenden Prüfern und konfrontierten diese mit den festgestellten Anker-Effekten. Die Prüfer gaben an, dass sie zu einem gewissen Ausmaß auf die ungeprüften Buchwerte vertrauten, weil aus der Aufgabenstellung keine Hinweise hervorgingen, die für ein erhöhtes Risiko und somit gegen die Zuverlässigkeit der Daten sprachen. Allerdings ist auch nicht auszuschließen, dass es sich hierbei um Vorwände handelt, um die Beeinflussbarkeit zu rechtfertigen.

<sup>4</sup> Hierbei handelt es sich um die Studien von KINNEY/UECKER (1982), BIGGS/WILD (1985), HEINTZ/WHITE (1989) und KENNEDY (1995).

<sup>5</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 135.

Mechanismus zu untersuchen, ließen die Autoren insgesamt 252 Prüfer mit einer durchschnittlichen Prüfungserfahrung von etwa sechs Jahren eine analytische Prüfungshandlung durchführen.<sup>1</sup> Dabei integrierten sie in ihr Untersuchungsdesign insgesamt drei mögliche Ausprägungen des Finanzbuchführungssystems des Mandanten (im Folgenden als *data environment* bezeichnet).<sup>2</sup> Zwei davon signalisierten potenzielle Probleme im *data environment* (Änderungen im Informationssystem im zu prüfenden Geschäftsjahr und wesentliche Nachbuchungen im Rahmen der Vorjahresprüfung). Eine weitere Bedingung enthielt keine Hinweise, die auf potenzielle Probleme im *data environment* schließen ließen (sog. *clean condition*). Die Autoren gingen davon aus, dass entsprechende Problemhinweise einem möglichen „Vertrauen“ der Prüfer in den ungeprüften Buchwert entscheidend entgegenwirken könnte.<sup>3</sup> Insgesamt basierte die Studie auf einem  $3 \times 2 \times 2$  Untersuchungsdesign (*data environment*  $\times$  Vorjahrestrend  $\times$  ungeprüfter Buchwert) mit den drei oben dargestellten Ausprägungen des *data environment*, zwei möglichen Trends in den Vorjahresdaten (steigend/fallend) und zwei potenziellen Richtungen des ungeprüften Buchwerts (steigend/fallend).<sup>4</sup>

Zunächst einmal ist festzuhalten, dass HEINTZ/WHITE/BEDARD erneut einen empirischen Nachweis dafür erbringen konnten, dass Prüfer signifikant von den ungeprüften Buchwerten beeinflusst wurden, wenn sie nicht explizit auf Probleme in dem *data environment* des Mandanten hingewiesen wurden. Insofern konnten die Ergebnisse der vorherigen Studien erneut bestätigt werden.<sup>5</sup> Allerdings ließen die Ergebnisse auch erkennen, dass das Ausmaß des Anker-Effekts durch Hinweise auf potenzielle Probleme im *data environment* des Mandanten in einigen, jedoch nicht in allen Fällen, erheblich reduziert werden konnte.<sup>6</sup> Ein *debiasing* des Anker-Effekts konnte immer beobachtet werden, wenn der Hinweis auf die mangelnde Zuverlässigkeit der Daten gemeinsam mit einem im Vergleich zum Vorjahr gestiegenen (ungeprüften) Buchwert auftrat. In den anderen Situationen ließen sich die teilnehmenden Prüfer trotz entsprechender Problemhinweise weiterhin von den ungeprüften Buchwerten beeinflussen.<sup>7</sup>

---

<sup>1</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 138. Da die Aufgabenstellung ansonsten analog zu HEINTZ/WHITE (1989) konzipiert war, wird hier nicht weiter darauf eingegangen. Vgl. hierzu ausführlich Kap. 3.5.2.2.2.1.

<sup>2</sup> Vgl. im Folgenden HEINTZ/WHITE/BEDARD (1999), S. 136.

<sup>3</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 139.

<sup>4</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 138-139.

<sup>5</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 140-141.

<sup>6</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 141.

<sup>7</sup> Vgl. HEINTZ/WHITE/BEDARD (1999), S. 143.



Möchte man abschließend die Ergebnisse der beiden o.a. *debiasing*-Studien ergänzend zu den bisherigen Studien heranziehen, so können folgende Erkenntnisse festgehalten werden:

- 4) Prüfer lassen sich unabhängig von der Anzahl der Berufsjahre von den ungeprüften Werten beeinflussen. Erfahrung scheint daher nicht dazu geeignet, um als *debiasing*-Mechanismus für das Vermeiden bzw. Verringern von Framing-Effekten bei der Durchführung von analytischen Prüfungshandlungen herangezogen werden zu können.
- 5) Prüfer lassen sich trotz erhöhter Verantwortung von den ungeprüften Werten beeinflussen. *Accountability* scheint daher ebenfalls nicht als *debiasing*-Mechanismus geeignet zu sein.
- 6) Prüfer lassen sich weniger von den ungeprüften Werten beeinflussen, wenn sie Gegenargumente für die Richtigkeit des ungeprüften Werts erbringen sollen. *Counterexplanations* scheinen daher als *debiasing*-Mechanismus zu funktionieren.
- 7) Prüfer lassen sich nicht durch einen ungeprüften Buchwert beeinflussen, wenn Hinweise existieren, die auf die Unzuverlässigkeit des ungeprüften Buchwerts hindeuten, und der ungeprüfte Buchwert zudem höher ist als der entsprechende Vorjahreswert. Erhöhtes Prüfungsrisiko wirkt dem Anker-Effekt entgegen und kann demnach als funktionierender *debiasing*-Mechanismus bezeichnet werden.

Die vierte Erkenntnis verdeutlicht die besondere Gefahr, die von einer möglichen Beeinflussung des Prüfers durch die Kenntnis des ungeprüften Buchwerts ausgeht. Analytische Prüfungshandlungen werden in der Prüfungspraxis zwar überwiegend von Prüfern vorgenommen, die bereits über ein Mindestmaß an Prüfungserfahrung verfügen. Prüfungsleiter, die durchschnittlich drei Jahre Berufserfahrung aufweisen, führen analytische Prüfungshandlungen regelmäßig im Rahmen der Planungsphase zur Identifikation von Prüfungsrisiken und zur Erstellung des Prüfungsplans durch. Da Prüfungsleiter in der Regel auch an der Phase der Prüfungsdurchführung beteiligt sind, ist denkbar, dass der den Prüfungsplan erstellende Prüfungsleiter Prüffelder wiederholt verprobt, um entweder die Eindrücke aus der Planungsphase zu bestätigen bzw. zu revidieren oder um neu gewonnene Erkenntnisse einfließen zu lassen. Manager, die bereits über erheblich mehr Prüfungserfahrung als Prüfungsleiter verfügen, sind in der Regel am Review-Prozess und an der abschließenden Gesamtdurchsicht beteiligt. Im Review-Prozess werden die Prüfungshandlungen, also auch die analytischen Prüfungshandlungen, des ge-

samen Prüfungsteams nachvollzogen und beurteilt. Im Rahmen der abschließenden Gesamtdurchsicht machen analytische Prüfungshandlungen einen großen Teil der Arbeit aus, da hier aufgrund der geringen zur Verfügung stehenden Zeit keine zeitintensiven Einzelfallprüfungen in Betracht kommen. Bei jungen unerfahrenen Prüfungsassistenten<sup>1</sup> (Berufseinsteiger) kommen analytische Prüfungshandlungen hingegen kaum zum Einsatz, da sie zumeist mit den einfach durchzuführenden, aber zeit- und kostenintensiven Einzelfallprüfungen in risikogeringen Prüffeldern beauftragt werden. Da aber Erfahrung kein *debiasing* auf den Anker-Effekt entfaltet, ist zu vermuten, dass die Kenntnis des ungeprüften Buchwerts die Effektivität und Effizienz von analytischen Prüfungshandlungen negativ beeinträchtigt.

Die fünfte Erkenntnis ist ebenfalls als kritisch einzuschätzen. *Accountability* ist ein dem Prüfungsprozess immanenter Mechanismus. Da Prüfer ihre Prüfungshandlungen und -urteile im Rahmen des Review-Prozesses regelmäßig rechtfertigen und verantworten müssen, ist *Accountability* im Rahmen der Abschlussprüfung zwar institutionalisiert. Ihre mangelnde Eignung als *debiasing*-Mechanismus lässt jedoch befürchten, dass mit Effektivitäts- und effizienzmindernden Anker-Effekten in der Prüfungspraxis zu rechnen ist.

Die sechste Erkenntnis zeigt, dass Anker-Effekte *effort-related* sind, die durch *effort-induced debiasing*-Mechanismen gemildert werden können. *Counterexplanations*, also die Aufforderung, sich die „Ungeprüftheit“ des Buchwerts und seiner potenziellen „Unrichtigkeit“ bewusst zu werden, können dazu führen, dass sich der Prüfer nicht mehr so stark von dem ungeprüften Buchwert beeinflussen lässt und die Effizienz und Effektivität von analytischen Prüfungshandlungen gesteigert oder zumindest nicht beeinträchtigt wird. Möglichkeiten zur Integration von *counterexplanations* in den Prüfungsprozess sind ebenfalls schnell auffindbar. Eine erste Möglichkeit ist natürlich in der Aufnahme einer entsprechenden Aufforderung in die Prüfungsstandards zu sehen. Darüber hinaus ist auch eine Berücksichtigung in der jeweiligen Prüfungssoftware der Prüfungsgesellschaften denkbar. Aufforderungen zur kritischen Beurteilung der ungeprüften Buchwerte des Mandanten könnten an unterschiedlichen Stellen des Prüfungsprozesses auftauchen. Etwa zu Beginn der Prüfungsplanung, wenn der Prüfer die vorläufigen Jahresabschlussdaten des Mandanten in seine Prüfungssoftware übernimmt, oder im Rahmen der Dokumentation der einzelnen Prüfungshandlungen als explizite Aufforderung, kritisch

---

<sup>1</sup> Der Prüfungsassistent steht auf der untersten Stufe der Mitarbeiterhierarchie einer Prüfungsgesellschaft. Vgl. MARTEN/QUICK/RUHNKE (2006), S. 594.

Stellung zu den ungeprüften Buchwerten zu nehmen. Weiterhin könnte man bereits sehr früh im Ausbildungsprozess, etwa in den Schulungsmaßnahmen der Prüfungsgesellschaften oder schon im Rahmen der universitären Ausbildung, damit beginnen, Prüfer für die Notwendigkeit von *counterexplanations* zu sensibilisieren.

Die siebte Erkenntnis weist darauf hin, dass Prüfer nicht oder nur geringfügig von einem ungeprüften Buchwert beeinflusst werden, falls Informationen darauf hinweisen, dass die Umsätze überbewertet sein könnten und darüber hinaus die Zuverlässigkeit der vorläufigen Jahresabschlussdaten des Mandanten erheblich in Frage gestellt ist. In der Studie von HEINTZ/WHITE/BEDARD wurde ein erhöhtes Risiko durch die Kombination von im Vergleich zum Vorjahr gestiegenen Umsatzerlösen und Mängeln im *data environment* des Mandanten signalisiert. Zweifelsohne ist dies eine Situation, in welcher eine erhöhte Wachsamkeit vom Prüfer besonders gefordert ist und Beeinflussungsmöglichkeiten von Seiten des Mandanten gänzlich abzuschalten sind. Insofern scheinen die Ergebnisse von HEINTZ/WHITE/BEDARD, die Gefahr effektivitäts- und effizienzmindernder Anker-Effekte bei analytischen Prüfungshandlungen erheblich zu entschärfen.<sup>1</sup> Dennoch sollte nicht gleich davon ausgegangen werden, dass die Kenntnis des ungeprüften Buchwerts völlig unbedenklich ist. Zu bedenken gibt etwa, dass die an der Studie von HEINTZ/WHITE/BEDARD teilnehmenden Prüfer explizit auf die Mängel in der *data environment* hingewiesen wurden. In der Praxis mögen Indikatoren für ein erhöhtes Risiko nicht immer so hervorstechend sein. Der Prüfer könnte sie schlicht und einfach nicht bemerken, insbesondere, wenn der Mandant zudem ein Interesse daran hat, sie zu verschleiern. In diesem Zusammenhang soll auch auf die in der jüngsten Zeit zunehmenden Fälle von Bilanzmanipulationen hingewiesen werden. Ein zum Mandanten durch langjährige Geschäftsbeziehungen aufgebautes Vertrauen könnte ebenso wie die Gefahr zur Betriebsblindheit dazu führen, dass auf Probleme deutende Signale nicht immer direkt erkannt werden.

Darüber hinaus sind auch noch diejenigen Konstellationen in der HEINTZ/WHITE/BEDARD-Studie zu beachten, die nicht zu einem *debiasing* des Anker-Effekts führten. So können Umsätze auch dann überbewertet sein, wenn de facto keine Probleme in der *data environment* vorliegen. Außerdem kann eine Überwertung der Umsätze auch dann vorliegen, wenn sie im Vergleich zum Vorjahr gesunken oder konstant geblieben sind. Da die Autoren bei derartigen Sachverhalten weiterhin Framing-

---

<sup>1</sup> So argumentieren auch HEINTZ/WHITE/BEDARD (1999), S. 137.

Effekte feststellen konnten, sollte die von den ungeprüften Buchwerten ausgehende Gefahr für die Effizienz und Effektivität von analytischen Prüfungshandlungen weiterhin ernst genommen werden.

Weiterhin kann argumentiert werden, dass die Aufgabe des Prüfers nicht ausschließlich in der Aufdeckung von Fehlern besteht, die zu Überbewertungen führen. Unterbewertungen der Umsatzerlöse könnten zu falschen Signalen bei Kapitalgebern, Gläubigern oder sonstigen Geschäftspartnern führen. Der Abzug von liquiden Mitteln aufgrund von schwindendem Vertrauen der Kapitalgeber kann ebenso zu Liquiditätsproblemen führen wie erhöhte Zinsen aufgrund von gestiegenen Risikozuschlägen. Derartige Folgen könnten den Beginn einer ernsthaften Gefährdung des Unternehmensbestands darstellen, welcher der Mandant durch eine (unverzerrte) Pflichterfüllung des Abschlussprüfers möglicherweise nicht gegenüber stünde.

Schließlich ist denkbar, dass Unterbewertungen auch durch systematische Fehler verursacht werden. Systematische Fehler könnten an anderer Stelle zu Überbewertungen führen. Insofern liefert ihre Aufdeckung im Rahmen von Unterbewertung Hinweise zur Identifikation von Überbewertungen. Daher kann nicht ernsthaft behauptet werden, dass Framing-Effekte bei im Vergleich zum Vorjahr gesunkenen Buchwerten unproblematisch sind. Zusammenfassend bleibt also zu konstatieren, dass Framing-Effekte als Gefährdung für die Effizienz und Effektivität von analytischen Prüfungshandlungen und somit den gesamten Prüfungsprozess weiterhin ernst genommen werden sollte.

### **3.5.2.2.3 Favourable Outcome-Effekt**

Eine weitere Möglichkeit der Verzerrung im Rahmen der Problemrepräsentation stellt der sog. *favourable outcome* effect dar. In der psychologischen Forschung konnte festgestellt werden, dass der Grad, zu dem ein erlangtes Ergebnis mit den Zielen bzw. den Wünschen des Individuums übereinstimmt, die Verlässlichkeitseinschätzung des Ergebnisses beeinflussen kann.<sup>1</sup> In Bezug auf die Durchführung von analytischen Prüfungshandlungen könnte sich eine derartige Verzerrung auf die Festlegung von Toleranzgrenzen auswirken. Wie in Kap. 2.3.2.2.2 gezeigt wurde, sollte der Prüfer bei der Festlegung von Toleranzgrenzen unter anderem die Genauigkeit des entwickelten Erwartungswerts berücksichtigen, welcher seinerseits maßgeblich von dem Aggregati-

---

<sup>1</sup> Vgl. ELSTER (1999); HASTIE (1984).

onsgrad der berücksichtigten Daten abhängt. Wäre die prüferische Urteilsbildung allerdings von dem favourable outcome effect gekennzeichnet, so bestünde die Gefahr, dass Prüfer bei der Festlegung von Toleranzgrenzen nicht auf die objektiven Kriterien zurückgreifen würden. Um diesen Aspekt zu untersuchen, ließen GLOVER/PRAWITT/WILKS 67 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 34,5 Monaten eine analytische Prüfungshandlung zur Beurteilung der Zinserträge durchführen, wobei der mitgeteilte Erwartungswert auf Basis von hoch aggregierten Daten (Jahreszahlen) entwickelt wurde. Dabei wurde einer Hälfte der Probanden ein Erwartungswert mitgeteilt, der sich wesentlich von dem zu prüfenden Buchwert unterschied (not favourable outcome group). Bei der anderen Hälfte wich der Erwartungswert hingegen nicht wesentlich vom Buchwert ab (favourable outcome group). Nachdem die Probanden die Ergebnisse zur Kenntnis genommen haben, wurden sie dazu aufgefordert, die Verlässlichkeit des Ergebnisses der analytischen Prüfungshandlung – mithin die Aussagekraft des erlangten Prüfungsnachweises – einzuschätzen. Anschließend wurde beiden Gruppen erneut ein Erwartungswert mitgeteilt, der allerdings auf Basis von weniger aggregierten Daten (Quartalsdaten) generiert wurde und in beiden Gruppen eine wesentliche Abweichung vom ungeprüften Buchwert indizierte. Hiernach wurden die Prüfer aufgefordert, sowohl die Verlässlichkeit der neuen Ergebnisse als auch der alten Ergebnisse einzuschätzen.<sup>1</sup>

Ingesamt konnten die Autoren einen signifikanten Zusammenhang zwischen dem Ergebnis der ersten analytischen Prüfungshandlung (favourable oder not favourable) und der Verlässlichkeitseinschätzung feststellen. Obwohl der Erwartungswert auf hoch aggregierten Daten basierte – und somit eine geringe Genauigkeit aufwies – schätzten Prüfer der favourable outcome Gruppe die Verlässlichkeit der Ergebnisse wesentlich höher ein als Prüfer der not favourable outcome Gruppe (3,9 ggü. 2,9).<sup>2</sup> Die Vermutung der Autoren, dass die Prüfer von einem favourable outcome effect betroffen sind, konnte mithin bestätigt werden. Zudem stellten die Autoren fest, dass die Verlässlichkeitseinschätzung des ersten Ergebnisses von den Prüfern der favourable outcome Gruppe wesentlich stärker reduziert wurde (von 3,9 auf 2,7) als von Prüfern in der not favourable outcome Gruppe (von 2,9 auf 2,8), nachdem ihnen der auf Basis von weniger aggregierten Daten entwickelte Erwartungswert mitgeteilt wurde.<sup>1</sup> Daraus schlussfolgerten die Autoren, dass diese Verzerrung nicht auf das fehlende Wissen der Prüfer zurückgeführt

---

<sup>1</sup> Vgl. GLOVER/PRAWITT/WILKS (2005), S. 205.

<sup>2</sup> Vgl. GLOVER/PRAWITT/WILKS (2005), S. 209.

werden konnte, dass hoch aggregierte Daten zu weniger verlässlichen Ergebnissen führen. Vielmehr war es den Prüfern nicht möglich, dieses Wissen abzurufen. Nachdem ihnen allerdings anhand des auf Quartalszahlen abgeleiteten Erwartungswerts ein Hinweis (prompt) auf die Ungenauigkeit des auf Jahreszahlen basierenden Erwartungswerts geliefert wurde, war es ihnen möglich, ihr Unvermögen zu überwinden und die Verlässlichkeitseinschätzung entsprechend des Aggregationsgrads der berücksichtigten Daten anzupassen.<sup>2</sup>

Um einem favourable outcome effect in der Prüfungspraxis entgegenzuwirken, scheint es also weniger zielführend, den Zusammenhang zwischen der Beweiskraft von Prüfungsnachweisen und dem Aggregationsgrad der herangezogenen Daten etwa im Rahmen von Schulungen, Seminaren oder Lehrbüchern zu betonen, da die Ergebnisse der Studie von GLOVER/PRAWITT/WILKS darauf hinweisen, dass nicht das fehlende Wissen für die Verzerrungen verantwortlich ist. Da IDW PS 312.21 sowie ISA 520.12e bereits ausdrücklich auf die Berücksichtigung des Aggregationsgrads von Daten, die zur Durchführung von analytischen Prüfungshandlungen herangezogen werden, hinweisen, ist auch dieses Mittel bereits ausgeschöpft. Eine probate Maßnahme zur Vermeidung des favourable outcome effects könnte jedoch darin bestehen, dass Prüfer im Rahmen des Prüfungsprozesses an die geringe Aussagekraft von auf Basis hochaggregierter Daten erlangten Prüfungsnachweisen „erinnert“ werden, so dass sie das potenziell vorhandene Wissen über diesen Zusammenhang entsprechend abrufen und im Rahmen ihres Urteilsbildungsprozesses berücksichtigen können. Derartige Hinweise (sog. prompts) könnten problemlos in den Prüfungsprozess integriert werden. Beispielsweise könnten Prüfer im Rahmen der Übernahme der ungeprüften Jahresabschlussdaten in die eigene Prüfungssoftware erstmalig auf die geringere Beweiskraft von hochaggregierten Daten hingewiesen werden. Eine weitere Möglichkeit, das Wissen über den Zusammenhang zwischen Beweiskraft und Aggregationsgrad zu aktivieren, besteht in Rahmen der Dokumentation von Prüfungsnachweisen. Auch hier könnte den Prüfern ein Hinweis erteilt werden, der ihnen den Zusammenhang erneut verdeutlicht.

---

<sup>1</sup> Vgl. GLOVER/PRAWITT/WILKS (2005), S. 209.

<sup>2</sup> Vgl. GLOVER/PRAWITT/WILKS (2005), S. 213.

### **3.5.2.3 Der Aufbau einer inadäquaten Problemrepräsentation durch den Erhalt einer Mandantenerklärung und der Wechsel zu einer adäquaten Problemrepräsentation**

Empirische Studien aus dem angloamerikanischen Sprachraum weisen darauf hin, dass Prüfer häufig den Mandanten aufsuchen, um eine Begründung für eine unerwartete Abweichung zu erhalten.<sup>1</sup> Darüber hinaus wird ein derartiges Vorgehen sogar in den Prüfungsstandards des Berufsstands empfohlen.<sup>2</sup> Die Einholung einer Mandantenerklärung kann sich grundsätzlich auch als sinnvoll erweisen, solange sie dem tatsächlichen Sachverhalt entspricht und den Prüfer gleich zu Beginn einer analytischen Prüfungshandlung zum Aufbau einer adäquaten Problemrepräsentation bewegt. In derartigen Fällen kann die Befragung des Mandanten sogar zu Effizienzvorteilen für den Prüfungsauftrag führen. Möglicherweise führen Mitarbeiter des Mandanten die festgestellten Abweichungen jedoch bewusst oder unbewusst auf Ursachen zurück, die nicht oder nur zum Teil der Wahrheit entsprechen. In solchen Fällen könnten Mandantenerklärungen dazu beitragen, dass der Prüfer zu Beginn seiner analytischen Prüfungshandlung eine inkorrekte Problemrepräsentation aufbaut. Der Erfolg der weiteren Problembearbeitung hängt in derartigen Fällen entscheidend davon ab, ob der Problemlöser seine Repräsentation wechseln kann. In der Problemlösungsforschung spricht man hierbei auch von Umstrukturierung- oder Neuentdeckung des Problemraums.<sup>3</sup>

In diesem Zusammenhang ließen BIERSTAKER/BEDARD/BIGGS zwölf Prüfungsleiter eine analytische Prüfungshandlung im Bereich des Vorratsvermögens durchführen.<sup>4</sup> Dabei ließen die Autoren den an der Studie teilnehmenden Prüfern eine Erklärung des Mandanten zukommen, die nicht den eigentlichen Grund für die festgestellten Diskrepanzen einnehmen konnte. Alle Prüfer starteten ihre Problembearbeitung mit einer „falschen“, auf die Mandantenerklärung bezogenen Problemrepräsentation. Lediglich einer von ihnen zeigte sich in der Lage, seine Problemrepräsentation selbstständig zu wechseln. Von den restlichen elf Prüfern waren zehn erst dann in der Lage, zu einer adäquaten Repräsentation zu wechseln, nachdem ihnen externe Hilfestellungen zur Verfügung gestellt wurden. Ein weiterer Prüfer war trotz weiterer Hilfestellungen nicht im Stande, seine anfängliche „falsche“ Repräsentation zu ändern.

---

<sup>1</sup> Vgl. etwa BEDARD/BIGGS (1991b), S. 77; HIRST/KOONCE (1996), S. 462, 471.

<sup>2</sup> Vgl. IDW PS 312.27 sowie ISA 520.18.

<sup>3</sup> Vgl. hierzu insbesondere Kap. 3.2.2.3.3.2.5.

<sup>4</sup> Da es sich um einen identischen Versuchsaufbau wie bei Bedard/Biggs (1991a) handelt, soll an dieser Stelle auf die entsprechenden Ausführungen unter Kap. 3.5.2.1 verwiesen werden.

BIERSTAKER/BEDARD/BIGGS stellten fest, dass der Wechsel der Problemrepräsentation gerade bei denjenigen Prüfern mit erheblichen Schwierigkeiten verbunden war, welche die Mandantenerklärung bereits sehr früh in ihrem Urteilsbildungsprozess berücksichtigt haben. Hiervon waren zehn der zwölf Prüfer betroffen. Einer der beiden übrigen Prüfer stützte sich erst in einer späteren Phase seiner Urteilsfindung auf die Mandantenerklärung und hatte anschließend weniger Probleme, einen Wechsel zu einer produktiveren Problemrepräsentation zu vollziehen. Daraus schlussfolgerten die Autoren, dass der Wechsel der Repräsentation umso problematischer ist, je eher die Mandantenerklärung in die prüferische Urteilsbildung einfließt.<sup>1</sup> Zusammenfassend konstatieren BIERSTAKER/BEDARD/BIGGS, dass der Erhalt einer Mandantenerklärung den Wechsel von einer unproduktiven zu einer produktiven Problemrepräsentation und somit den gesamten Erfolg bei der Durchführung von analytischen Prüfungshandlungen erheblich beeinträchtigt.<sup>2</sup>

Die Erkenntnisse von BIERSTAKER/BEDARD/BIGGS lassen erhebliche Zweifel an der Zweckmäßigkeit der Prüfungsstandards aufkommen. IDW PS 312.27 sowie ISA 520.18 fordern den Prüfer explizit dazu auf, bei festgestellten Abweichungen zu allererst mit der Befragung des Mandanten zu beginnen. Eine derartige Vorgehensweise kann jedoch nur befürwortet werden, wenn der Mandant den Grund für die Abweichung tatsächlich kennt und „unverzerrt“ an den Prüfer weitergibt. Jedoch sollte hiervon nicht grundsätzlich ausgegangen werden. Vorstellbar wäre etwa, dass der Mandant aufgrund von Zeitdruck unabsichtlich eine falsche Erklärung abgibt, die zwar dem ersten Anschein nach die Diskrepanzen erklären könnte. Darüber hinaus ist auch nicht von der Hand zu weisen, dass Mandanten dem Prüfer bewusst eine irreführende Information weitergeben könnten. Hierfür sprechen auch die zahlreichen in der jüngeren Vergangenheit bekannt gewordenen Fälle von Bilanzmanipulationen.

Sollten sich also Prüfer in der Praxis (in der Studie waren es insbesondere Prüfungsleiter) tatsächlich entsprechend dem Wortlaut der Prüfungsstandards verhalten und eine inkorrekte Mandantenerklärung erhalten, besteht die Gefahr, dass er diese bereits sehr früh in seine Problemrepräsentation integriert. Dieses könnte eine erhebliche Gefahrenquelle für die Effektivität und Effizienz von Jahresabschlussprüfungen darstellen. Eine Möglichkeit, diesem Risiko entgegenzuwirken, ist die Modifikation des Wortlauts der Standards oder zumindest die entsprechende Aufnahme eines Hinweises in den Prü-

---

<sup>1</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 31-32.

<sup>2</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 33 i.V.m. 35.



fungshandbüchern der Prüfungsgesellschaften. Prüfer sollten darauf aufmerksam gemacht werden, dass sie bei der Feststellung unerwarteter Abweichungen nicht gleich den Mandanten konsultieren sollten. Vielmehr sollten sie sich zunächst eine auf eigenen Erklärungen basierende Problemrepräsentation aufbauen. Die Mandantenerklärung sollte nur alternativ bzw. im Anschluss an eine eigenständige Hypothesengenerierung eingeholt werden. Nur auf diesem Wege kann der Prüfer sicherstellen, dass er seine kognitive Flexibilität nicht verliert.

### 3.5.2.4 Tabellarische Darstellung der empirischen Befunde zur Phase der Problemrepräsentation

| Autor(en)                          | Teilnehmer |                                                           | Daten-<br>gewinnung | Daten-<br>auswertung | Ergebnisse                                                                                                                                                                                              | Signifikanz |
|------------------------------------|------------|-----------------------------------------------------------|---------------------|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
|                                    | Anzahl     | Art                                                       |                     |                      |                                                                                                                                                                                                         |             |
| BIGGS/MOCK/WATKINS<br>(1988)       | 4          | 2 Manager<br>Ø = 7 J.<br><br>2 Prüfungsleiter<br>Ø > 3 J. | Experiment          | Verbale Protokolle   | Manager benötigen weniger Zeit zum Aufbau einer adäquaten Problemrepräsentation, richten ihre Aufmerksamkeit weniger auf oberflächliche Informationen und treffen die richtigen Prüfungsentscheidungen. | k.A.        |
| BEDARD/BIGGS<br>(1991a)            | 21         | 10 Manager<br>k.A.<br><br>11 Prüfungsleiter<br>k.A.       | Experiment          | Verbale Protokolle   | Manager sind vermehrt dazu in der Lage, ein zusammenhängendes Muster zu erkennen (sog. Mustererkennung).                                                                                                | k.A.        |
| BIERSTAKER/BEDARD/<br>BIGGS (1999) | 12         | Prüfungsleiter<br>3 bis 5 J.                              | Experiment          | Verbale Protokolle   | Prüfungsleiter sind erst nach externen Hilfestellungen (sog. prompts) dazu in der Lage, ein zusammenhängendes Muster zu erkennen.                                                                       | k.A.        |
| AYERS/KAPLAN (1993)                | 45         | Prüfungsleiter<br>Ø = 3,3 J.                              | Experiment          | Kovarianzanalyse     | Prüfer der misstatement-frame-Gruppe listen signifikant mehr Fehler-Hinweise auf                                                                                                                        | 0,065       |
|                                    |            |                                                           |                     | Kovarianzanalyse     | Prüfer der non-misstatement-frame-Gruppe listen signifikant mehr Nicht-                                                                                                                                 | 0,024       |

|                      |     |                                                                      |            |                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                      |
|----------------------|-----|----------------------------------------------------------------------|------------|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
|                      |     |                                                                      |            |                                                | Fehler-Hinweise auf.                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                      |
| KINNEY/UECKER (1982) | 179 | Prüfungsleiter k.A.                                                  | Experiment | Chi-Quadrat-Test/<br>t-Test                    | Prüfer mit einem niedrigen Buchwert setzten geringere investigation boundaries als Prüfer mit einem hohen Buchwert.                                                                                                                                                                                                                                                                                                                                             | UG: $<0,03^1$ / $<0,01^2$<br>OG: $<0,09$ / $<0,01$                                                                   |
| BIGGS/WILD (1985)    | 121 | Prüfungsassistenten, Prüfungsleiter, Manager, Partner<br>1 bis 24 J. | Experiment | Mann-Whitney-U-Test                            | Prüfer werden von einem hohen Buchwert i.V.m. einer Trendwende stärker beeinflusst als von einem niedrigen Buchwert i.V.m. mit einer Trendfortführung.                                                                                                                                                                                                                                                                                                          | 0,001/0,015/0,003 (EW/OG/UG)<br>ggü.<br>0,107/0,027/0,353 (EW/OG/UG) <sup>3</sup>                                    |
| HEINTZ/WHITE (1989)  | 138 | k.A.<br>Ø = 5,5 J.                                                   | Experiment | Mann_Whitney-U-Test<br><br>Mann_Whitney-U-Test | Prüfer werden leicht stärker von einem ungeprüften Buchwert beeinflusst, der einen fallenden Trend fortführt, als von einem ungeprüften Buchwert, der einen steigenden Trend fortführt.<br><br>Prüfer werden wesentlich stärker von einem ungeprüften Buchwert beeinflusst, der eine Trendwende darstellt, als von einem ungeprüften Buchwert, der eine Trendfortführung darstellt; unabhängig davon, ob ein sinkender oder steigender Trend durchbrochen wird. | 0,00 <sup>1</sup><br><br>0,00 <sup>2</sup><br>Trendwende nach oben<br><br>0,00 <sup>3</sup><br>Trendwende nach unten |

<sup>1</sup> Signifikanzniveau des Chi-Quadrat-Tests.

<sup>2</sup> Signifikanzniveau des t-Tests.

<sup>3</sup> Obwohl der Vergleich von Signifikanzniveaus keine direkten Rückschlüsse über die unterschiedliche Stärke eines Einflusses ermöglicht, werten BIGGS/WILD diese Ergebnisse als einen indirekten Nachweis für den stärkeren Einfluss eines ungeprüften hohen Buchwerts i.V.m. einer Trendwende im Vergleich zu einem ungeprüften niedrigen Buchwert i.V.m. einer Trendfortführung. Vgl. BIGGS/WILD (1985), S. 617.

|                             |     |                                      |            |                |                                                                                                                                                                                                          |         |
|-----------------------------|-----|--------------------------------------|------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| KENNEDY (1995)              | 551 | 229 Studenten<br>322 Manager<br>k.A. | Experiment | Varianzanalyse | Studenten und Manager lassen sich gleichermaßen durch einen ungeprüften Buchwert beeinflussen; d.h. Anker-Effekte treten unabhängig vom Umfang der Prüfungserfahrung auf.                                | > 0,1   |
|                             |     |                                      |            | Varianzanalyse | Verantwortungsdruck wirkt dem Anker-Effekt nicht entgegen und stellt somit keinen debiasing-Mechanismus dar.                                                                                             | > 0,1   |
|                             |     |                                      |            | Varianzanalyse | Der Einfluss von ungeprüften Buchwerten wird durch das Aufführen von Gegenargumenten (sog. Counterexplanations) gemildert. Counterexplanations stellen somit einen geeigneten debiasing-Mechanismus dar. | < 0,05  |
| HEINTZ/WHITE/BEDARD (1999)  | 252 | k.A.<br>Ø = 6 J.                     | Experiment | Varianzanalyse | Der Einfluss des ungeprüften Buchwerts wird gemildert, wenn ein Hinweis auf die mangelnde Zuverlässigkeit vorliegt und ein im Vgl. zum Vorjahr gestiegener ungeprüfter Buchwert vorliegt.                | < 0,05  |
| GLOVER/PRAWITT/WILKS (2005) | 67  | k.A.<br>Ø = 34,5 Mon.                | Experiment | Varianzanalyse | Prüfer schätzen die Verlässlichkeit von weniger zuverlässigen Ergebnissen höher ein, wenn das Ergebnis                                                                                                   | < 0,001 |

<sup>1</sup> Diese Angaben beziehen sich auf die Unterschiede bezüglich des Mittelwerts der angegebenen Erwartungswerte. Aus Gründen der Übersichtlichkeit wurde auf Angaben in Bezug auf die obere und untere Grenze verzichtet. Vgl. hierzu ausführlich HEINTZ/WHITE (1989), S. 32.

<sup>2</sup> Ebd.

<sup>3</sup> Ebd.

|                                    |    |                              |            |                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                               |
|------------------------------------|----|------------------------------|------------|---------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
|                                    |    |                              |            | <p>Varianzanalyse</p> <p>Varianzanalyse</p> | <p>keine Abweichung erkennen lässt (sog. favourable outcome Effekt). Prüfer, die vom favourable outcome Effekt betroffen sind, nehmen eine signifikante Änderung ihrer Verlässlichkeitseinschätzung vor, wenn sie anschließend Ergebnisse mit einem höheren Verlässlichkeitsgrad erhalten.</p> <p>Prüfer, die nicht vom favourable outcome Effekt betroffen sind, nehmen keine signifikante Änderung ihrer Verlässlichkeitseinschätzung vor, wenn sie anschließend Ergebnisse mit einem höheren Verlässlichkeitsgrad erhalten.</p> | <p>&lt; 0,001</p> <p>0,81</p> |
| BIERSTAKER/BEDARD/<br>BIGGS (1999) | 12 | Prüfungsleiter<br>3 bis 5 J. | Experiment | Verbale Protokolle                          | Prüfungsleiter, die eine Problemrepräsentation auf Basis der inkorrekten Mandantenerklärung aufbauen, haben Probleme, zu einer adäquaten Problemrepräsentation zu wechseln. Sie sind erst nach externen Hilfestellungen (sog. prompts) dazu in der Lage, zu einer adäquaten Problemrepräsentation zu wechseln.                                                                                                                                                                                                                     | k.A.                          |

Tabelle 1: Ergebnisse der empirischen Prüfungsforschung zur Phase der Problemrepräsentation

### **3.5.3 Studien zur Phase der Hypothesengenerierung**

#### **3.5.3.1 Vorbemerkungen**

Die Phase der Hypothesengenerierung nimmt eine entscheidende Bedeutung für die weitere prüferische Vorgehensweise ein. Sie steuert in einem durch Komplexität und Intransparenz geprägten Problemsachverhalt die Suche nach weiteren Informationen, um die zu Beginn noch unzulängliche Problemrepräsentation zu vervollständigen bzw. zu ergänzen. Im Rahmen von analytischen Prüfungshandlungen ist die Identifikation der Abweichungsursache maßgeblich von einer zweckmäßigen Hypothesengenerierung abhängig.<sup>1</sup>

Bei der Hypothesengenerierung kann einerseits nach dem Zeitpunkt bzw. Zeitraum der Generierung und andererseits nach der Quelle der generierten Hypothese differenziert werden. Im ersten Fall kann in eine anfängliche Hypothesengenerierung (sog. initial hypothesis generation) und eine nachfolgende Hypothesengenerierung (sog. subsequent hypothesis generation) klassifiziert werden.<sup>2</sup> Die anfängliche Hypothesengenerierung beginnt unmittelbar, nachdem der Prüfer eine unerwartete wesentliche Abweichung festgestellt hat, quasi im Anschluss an die Problemrepräsentation. Die nachfolgende Hypothesengenerierung findet hingegen erst im weiteren Verlauf der prüferischen Aktivitäten statt. In der Regel generiert der Prüfer neue Hypothesen, wenn er durch die Beschaffung von Informationen einen neuen Kenntnisstand erlangt hat.

Bezogen auf die Quelle der Hypothese ist zwischen der internen und der externen Hypothesengenerierung zu unterscheiden.<sup>3</sup> Interne Hypothesengenerierung bedeutet, dass der Prüfer einen potenziellen Abweichungsgrund aus dem Gedächtnis abrufen. Hierbei kommt dem Wissen des Individuums eine entscheidende Rolle zu. Bei der externen Hypothesengenerierung hingegen wird dem Prüfer eine potenzielle Abweichungsursache von außen, also der Prüfungsumwelt, zugetragen. Als unterschiedliche Quellen kommen hierbei der Mandant bzw. seine Mitarbeiter, der Vorgesetzte (i.d.R. handelt es sich hierbei um den Manager) sowie softwaregestützte Entscheidungshilfen, die i.d.R. in der Prüfungssoftware der Prüfungsgesellschaften integriert sind, in Frage.

---

<sup>1</sup> Vgl. BEDARD/BIGGS (1991a), S. 622.

<sup>2</sup> Vgl. etwa GREEN/TROTMAN (2003), S. 222.

<sup>3</sup> Siehe hierzu auch Kap. 2.3.2.4.2.

### 3.5.3.2 Unterteilung nach dem Zeitpunkt/-raum der Hypothesengenerierung

Die Differenzierung in eine anfängliche und nachfolgende Hypothesengenerierung betont den iterativen Charakter des Ablaufmodells zur Durchführung analytischer Prüfungshandlungen und berücksichtigt, dass Prüfer in einer fortgeschrittenen Phase ihres weiteren Vorgehens zusätzliche Informationen erlangen können, die sie zur Generierung weiterer Hypothesen veranlassen. Allerdings ist fraglich, ob Prüfer dazu in der Lage sind, eine inadäquate anfängliche Hypothesengenerierung (d.h. einen Hypothesensatz, der den tatsächlichen Grund nicht umfasst) durch die anschließende Generierung der korrekten Abweichungsursache zu kompensieren.<sup>1</sup> Gegen eine derartige Annahme sprechen frühe empirische Ergebnisse aus der medizinischen Forschung. PATEL/GROEN stellten fest, dass es den Probanden ausschließlich dann möglich war, korrekt zu diagnostizieren, wenn sie von Beginn an die richtige Hypothese aufstellten. Probanden, deren anfängliche Hypothesengenerierung den tatsächlichen Grund nicht umfasste, waren nicht in der Lage, die richtige Diagnose zu stellen.<sup>2</sup> Erste Indizien, die darauf hindeuteten, dass im Prüfungsbereich ein ähnlicher Sachverhalt zugrunde liegen könnte, führten BEDARD/BIGGS an.<sup>3</sup> Sie stellten fest, dass es ihren Probanden nicht möglich war, die richtige Abweichungshypothese nachträglich zu generieren, obwohl bereits entsprechende Informationen vorlagen.<sup>4</sup> Ähnliche Befunde erhielten BIERSTAKER/BEDARD/BIGGS.<sup>5</sup> Auch hier konnte festgestellt werden, dass die Probanden Schwierigkeiten hatten, trotz einer umfassenden Informationslage nachträglich die richtige Hypothese zu generieren.<sup>6</sup>

Um diesen Zusammenhang genauer zu untersuchen, ließen GREEN/TROTMAN 82 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,8 Jahren eine Fallstudie bearbeiten, bei der eine fehlerhafte Gemeinkostenzuordnung zu einer wesentlichen unerwarteten Abweichung in der Bruttogewinnmarge führte.<sup>7</sup> Nachdem die Prüfer dazu aufgefordert wurden, Hypothesen über die Ursache der Abweichung aufzustellen (anfängliche Hypothesengenerierung), konnten sie unterschiedliche Prüfungstests durchführen, deren

---

<sup>1</sup> So konnten ASARE/WRIGHT (2003), S. 248 feststellen, dass sich Prüfer trotz der Erlangung von Prüfungsnachweisen, die auf andere als die anfänglich generierten Abweichungshypothesen hindeuteten, nicht veranlasst sahen, eine nachfolgende Hypothesengenerierung durchzuführen. Da sich diese Studie allerdings auf die Phase der Informationssuche konzentrierte, soll sie an dieser Stelle nicht weiter vertieft werden. Eine genaue Darstellung der Studie sowie Analyse der Ergebnisse erfolgt in Kap. 3.5.4.3.

<sup>2</sup> Vgl. PATEL/GROEN (1986), S. 91-116.

<sup>3</sup> Eine ausführliche Beschreibung dieser Untersuchung findet sich in Kap. 3.5.2.1.

<sup>4</sup> Vgl. BEDARD/BIGGS (1991a), S. 640.

<sup>5</sup> Für eine ausführliche Fallbeschreibung siehe Kap. 3.5.2.3 i.V.m. 3.5.2.1.

<sup>6</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 29.

<sup>7</sup> Vgl. GREEN/TROTMAN (2003), S. 224-225.

Ergebnisse ihnen im Anschluss mitgeteilt wurden. Anschließend erhielten die Prüfer die Möglichkeit, weitere Abweichungshypothesen zu generieren (nachträgliche Hypothesengenerierung).<sup>1</sup>

Im Ergebnis stellten die Autoren fest, dass insgesamt 63 Prüfer die richtige Hypothese generierten. 50 der 63 Prüfer generierten den richtigen Abweichungsgrund in der anfänglichen Hypothesengenerierung und lediglich 13 in der nachträglichen Hypothesengenerierung.<sup>2</sup> Zwar zeigt dies, dass die Generierung der tatsächlichen Fehlerursache überwiegend in der anfänglichen Hypothesengenerierung erfolgt und es wesentlich seltener vorkommt, dass die richtige Ursache erst durch die Erlangung weiterer Informationen generiert wird. Dennoch weisen die Ergebnisse darauf hin, dass eine nachträgliche Generierung der richtigen Fehlerhypothese generell möglich ist. Da die Häufigkeit, mit welcher die richtige Hypothese nachträglich generiert wurde ( $n=13$ ), signifikant größer als null ist, kann konstatiert werden, dass die anfängliche Generierung der richtigen Fehlerursache keine notwendige Bedingung für die erfolgreiche Durchführung einer analytischen Prüfungshandlung ist.<sup>3</sup> Zudem stellten die Autoren fest, dass 39 der 50 Prüfer, deren anfängliche Hypothesengenerierung den tatsächlichen Abweichungsgrund beinhaltete, nicht dazu im Stande waren, die analytische Prüfungshandlung erfolgreich abzuschließen. Da auch diese Häufigkeit ( $n=39$ ) signifikant größer als null ist, konstatierten GREEN/TROTMAN, dass eine adäquate anfängliche Hypothesengenerierung weder notwendige noch hinreichende Bedingung für ein erfolgreiches prüferisches Vorgehen sei.<sup>4</sup>

Da die Erfolgsquote derjenigen Prüfer, die die richtige Hypothese erst im Nachhinein generierten, genauso hoch war, wie die Erfolgsquote derjenigen Prüfer, die die richtige Hypothese bereits zu Beginn generierten (38,5% gegenüber 38%), lässt die Untersuchung von GREEN/TROTMAN erkennen, dass eine iterative Vorgehensweise einen positiven Beitrag zur Erhöhung der Effektivität analytischer Prüfungshandlungen leisten kann.

Hierfür sprechen auch die Ergebnisse einer weiteren Studie von GREEN.<sup>1</sup> Sie ließ 58 Prüfer mit einer durchschnittlichen Prüfungserfahrung von vier Jahren die gleiche Aufgabe wie GREEN/TROTMAN bearbeiten und stellte dabei fest, dass knapp 64% der erfolg-

---

<sup>1</sup> Vgl. GREEN/TROTMAN (2003), S. 225-226.

<sup>2</sup> Vgl. GREEN/TROTMAN (2003), S. 227.

<sup>3</sup> Vgl. GREEN/TROTMAN (2003), S. 233.

<sup>4</sup> Vgl. GREEN/TROTMAN (2003), S. 227.



reichen Prüfer die richtige Fehlerhypothese erst im Laufe der Fallbearbeitung generierten.<sup>2</sup>

Noch deutlicher sind die Ergebnisse von ASARE/WRIGHT.<sup>3</sup> Sie stellten fest, dass alle Prüfer, die die richtige Fehlerhypothese erst im Nachhinein, d.h. nachdem sie weitere Informationen erlangt hatten, generierten, auch in der Lage waren, die richtige Fehlerursache zu identifizieren.<sup>4</sup>

Für die in der Praxis tätigen Prüfer ist es demnach wichtig zu erkennen, dass eine fortlaufende, sich der aktuellen Informationslage anpassende Hypothesengenerierung von besonderer Bedeutung ist. „(...) being able to ‚think outside the current hypothesis set‘“<sup>5</sup> kann demnach als ein wichtiger Faktor für die erfolgreiche Durchführung einer analytischen Prüfungshandlung aufgefasst werden.

### **3.5.3.3 Unterteilung nach der Quelle der Hypothese**

#### **3.5.3.3.1 Interne Hypothesengenerierung**

##### **3.5.3.3.1.1 Einflussfaktoren der ersten Hypothesengenerierung**

###### **3.5.3.3.1.1.1 Frequency knowledge**

Eine der zentralen Erkenntnisse der psychologischen Forschung ist darin zu sehen, dass die Wiederholung von bestimmten Sachverhalten bzw. die Häufigkeit von bestimmten Erfahrungen einen maßgeblichen Einfluss auf dessen nachfolgenden Abruf ausübt und das Lernen von Erscheinungshäufigkeiten (*frequency of occurrence*) mit sich bringt.<sup>1</sup> Im Kontext der Durchführung von analytischen Prüfungshandlungen spielt dieses als *frequency knowledge* bezeichnete Phänomen eine wesentliche Rolle bei der Generierung der ersten Hypothesen über den Grund einer festgestellten wesentlichen Abweichung. Dabei wird vermutet, dass die generierten Hypothesen eines Prüfers mehr Fehler umfassen, denen er im Rahmen seiner prüferischen Tätigkeit bereits häufiger begegnet ist

---

<sup>1</sup> Eine ausführliche Beschreibung der Untersuchung findet sich in Kap. 3.5.3.3.2.1.2.1.

<sup>2</sup> Vgl. GREEN (2004b), S. 15. Darüber hinaus stellte GREEN fest, dass es sich bei diesen Prüfern ausschließlich um Branchenspezialisten handelte. Da der Einfluss von branchenspezifischer Erfahrung jedoch noch ausführlich in Kap. 3.5.4.4.3.2 dargestellt wird, soll an dieser Stelle von einer Vertiefung dieses Aspekts abgesehen werden.

<sup>3</sup> Da diese Studie allerdings überwiegend die Phase der Informationssuche betrachtete, soll sie hier nicht weiter vertieft werden. Siehe hierzu ausführlich Kap. 3.5.4.3.

<sup>4</sup> Vgl. ASARE/WRIGHT (2003), S. 246.

<sup>5</sup> ASARE/WRIGHT (2003), S. 246.

(hohe Erscheinungshäufigkeit) als Fehler, denen er bisher seltener begegnet ist (geringe Erscheinungshäufigkeit).

Eine der ersten Studien, die diesen Sachverhalt empirisch untersuchte, führte LIBBY durch. Er forderte insgesamt 37 Manager mit einer Prüfungserfahrung zwischen drei und sechs Jahren (durchschnittlich fünf Jahre) auf, eine analytische Prüfungshandlung im Rahmen der Prüfungsplanung durchzuführen.<sup>2</sup> Dabei wurden den teilnehmenden Prüfern u.a. jeweils drei Kennzahlen (Bruttogewinnmarge, zweiter und dritter Liquiditätsgrad) aus dem zu prüfenden Geschäftsjahr und dem Vorjahr zur Verfügung gestellt. Die Kennzahlen des aktuellen Geschäftsjahrs beinhalteten einen bestimmten Fehler<sup>3</sup> und wiesen eine wesentliche Abweichung zu denen des Vorjahrs auf.<sup>4</sup> Die Probanden wurden aufgefordert, für die festgestellte Abweichung in Frage kommende Fehler aufzulisten. Eine Analyse der Antworten ergab, dass ein starker Zusammenhang zwischen den generierten Hypothesen und den Häufigkeiten, mit denen sich bestimmte Fehler im Jahresabschluss niederschlagen, existierte.<sup>5</sup>

LIBBY/FREDERICK erweiterten die Studie, indem sie das Untersuchungsdesign um unterschiedliche Umfänge an Berufserfahrung erweiterten. Im Gegensatz zu LIBBY ließen sie nicht ausschließlich Manager, sondern auch Prüfungsassistenten an ihrem Versuch teilnehmen. Die Manager wiesen einen Median von fünf Jahren Prüfungserfahrung auf, die Prüfungsassistenten einen Median von einem Jahr.<sup>6</sup> LIBBY/FREDERICK konnten zunächst feststellen, dass alle teilnehmenden Prüfer überwiegend Fehlerhypothesen mit einer hohen Eintrittshäufigkeit generierten. Insofern konnten die Ergebnisse von LIBBY dahingehend erweitert werden, dass nicht nur erfahrene Prüfer (Manager), sondern auch unerfahrene Prüfer (Prüfungsassistenten) über frequency knowledge verfügen. Allerdings konnten die Autoren weiterhin feststellen, dass erfahrene Prüfer (Manager) signifikant mehr Fehlerursachen mit einer hohen Eintrittshäufigkeit generierten als unerfahrene Prüfer (Prüfungsassistenten).<sup>7</sup>

---

<sup>1</sup> Vgl. KAHNEMAN/TVERSKY (1982), S. 143-157.

<sup>2</sup> Vgl. LIBBY (1985), S. 655.

<sup>3</sup> Es handelte sich dabei um einen nicht erfassten Einkauf von Vorräten im Wert von 2,5 Mio. \$, was etwa 5% des Gesamtvermögens ausmachte.

<sup>4</sup> Vgl. LIBBY (1985), S. 653.

<sup>5</sup> Vgl. LIBBY (1985), S. 657. Einen ähnlich signifikanten Zusammenhang konnten auch HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 771 feststellen.

<sup>6</sup> Vgl. LIBBY/FREDERICK (1990), S. 354. Darüber hinaus berücksichtigten die Autoren auch Studenten ohne Prüfungserfahrung.

<sup>7</sup> Vgl. LIBBY/FREDERICK (1990), S. 359.

Geht man davon aus, dass Fehler, die in der Vergangenheit von einer hohen Eintrittshäufigkeit geprägt waren, unter sonst gleichen Bedingungen mit einer hohen Wahrscheinlichkeit zukünftig der Grund für eine im Rahmen von analytischen Prüfungshandlungen festgestellte wesentliche Abweichung sind, so ist davon auszugehen, dass sich frequency knowledge positiv auf die Effektivität und Effizienz von analytischen Prüfungshandlungen auswirkt.<sup>1</sup> Sind die wahrscheinlichsten Fehlerursachen hingegen nicht Bestandteil eines vom Prüfer generierten Hypothesenbündels, so besteht die Gefahr, dass weniger wahrscheinliche Fehler als Grund für die Abweichung anerkannt werden, was tendenzielle Einbußen der Effektivität mit sich bringen könnte.<sup>2</sup>

Für die Prüfungspraxis würden diese Schlussfolgerungen implizieren, dass analytische Prüfungshandlungen im Rahmen der Prüfungsplanung tendenziell von erfahreneren Prüfern (etwa Managern) durchgeführt werden sollten.<sup>3</sup> Da eine derartige Personalzuordnung jedoch höhere Kosten zur Folge hätte, würde das Ziel einer wirtschaftlicheren Jahresabschlussprüfung jedoch verfehlt werden. Demnach gilt es, Wege zu finden, die eine Aneignung von frequency knowledge beschleunigen. Eine Möglichkeit wäre darin zu sehen, dass Prüfer nicht erst durch langjährige Prüfungserfahrung frequency knowledge erwerben, sondern eine Lernumgebung geschaffen wird, die den Erwerb von frequency knowledge bereits frühzeitig begünstigt. Dieses könnte etwa dadurch realisiert werden, dass bereits in Lehrbüchern, universitären Ausbildungsgängen mit Prüfungsbezug oder auch in Schulungsmaßnahmen von Prüfungsgesellschaften auf Fehlerhäufigkeiten, etwa in tabellarischer oder graphischer Form, hingewiesen wird.<sup>4</sup> Universitäten wären hierbei jedoch auf die Unterstützung der Prüfungspraxis angewiesen, da lediglich den Prüfungsgesellschaften Daten zu Fehlerhäufigkeiten aus der Prüfungsrealität zur Verfügung stehen. Dieses wäre sicherlich eine Möglichkeit, den häufig heraufbeschwörten Unterschied zwischen Theorie und Praxis zu überwinden und den Studenten bereits in einer sehr frühen Phase ihrer Ausbildung praxisrelevantes Wissen zur Verfügung zu stellen. Mittelfristig betrachtet würden sicherlich auch die Prüfungsgesellschaften von dieser Arbeitsteilung profitieren, da diese Studenten nach Studienabschluss möglicherweise den Einstieg in eine Prüfungsgesellschaft suchen.

---

<sup>1</sup> Vgl. LIBBY/FREDERICK (1990), S. 351-352, 363 sowie NELSON (1993), S. 806.

<sup>2</sup> Vgl. NELSON (1993), S. 806.

<sup>3</sup> Für Gründe, weshalb die Prüfungsplanung einen maßgeblichen Einfluss auf die Aufdeckung von wesentlichen Fehlern hat, siehe die Ausführungen unter Kap. 2.1.2.2, insbesondere Kap. 2.1.2.2.1.2.

<sup>4</sup> Vgl. hierzu BUTT (1988), S. 315-330; NELSON (1994), S. 185-211.

### 3.5.3.3.1.1.2 Recency effects

Ein weiterer Faktor, der die erste Hypothesengenerierung beeinflussen könnte, ist in der zeitlichen Reihenfolge zu sehen, mit der ein Prüfer bestimmte Fehler in der Vergangenheit aufgedeckt hat. Psychologische Studien weisen darauf hin, dass erst kürzlich erlebte Sachverhalte einfacher abgerufen (generiert) werden können als Erlebnisse, die bereits länger zurückliegen (sog. recency effect).<sup>1</sup> Im Rahmen der Durchführung von analytischen Prüfungshandlungen könnte dieses bedeuten, dass Fehler, die in der jüngeren Vergangenheit aufgedeckt wurden, mit einer höheren Wahrscheinlichkeit generiert werden als Fehler mit einem größeren zeitlichen Abstand. Um zu überprüfen, ob ein derartiger Zusammenhang in der Jahresabschlussprüfung zu beobachten ist, forderte LIBBY die 37 an seiner Studie teilnehmenden Manager nach Durchführung einer analytischen Prüfungshandlung dazu auf, die drei im Rahmen ihrer Prüfungstätigkeit zuletzt aufgedeckten Fehler aufzulisten.<sup>2</sup> Die Ergebnisse zeigten einen positiven Zusammenhang zwischen den erst kürzlich entdeckten Fehlern und den im Rahmen der Fallstudie generierten Hypothesen. Die Wahrscheinlichkeit, dass ein erst kürzlich entdeckter Jahresabschlussfehler generiert wurde, war anderthalb mal höher als die eines Fehlers, der nicht zu den drei kürzlich entdeckten Fehlern gehörte.<sup>3</sup> Somit konnte LIBBY den sog. recency effect in seiner Studie bestätigen.

Für die Prüfungspraxis ist ein mögliches Vorliegen des *recency effects* eher kritisch zu beurteilen, da ein derartiger Zusammenhang rational nicht begründet werden kann und das Prüfungsurteil somit subjektiv verzerrt wird. Da es sich bei den teilnehmenden Prüfern um Manager handelte, kann Erfahrung als möglicher *debiasing*-Mechanismus für die vorliegende Studie ausgeschlossen werden. Studien, die ein entsprechendes Untersuchungsdesign um accountability als möglichen debiaser erweitern, sind dem Autor bis zum heutigen Zeitpunkt nicht bekannt. Um diesen potenziellen Verzerrungen der prüferischen Urteilsbildung dennoch entgegenwirken zu können, sollten Prüfer darauf hingewiesen werden, dass kürzlich aufgedeckte Fehler nicht zwangsläufig auch der Grund für aktuell festgestellte Abweichungen sein müssen. Derartige Hinweise könnten mühelos in die Software zur Prüfungsdokumentation integriert werden. Darüber hinaus scheint es sinnvoll, Prüfer im Rahmen von Fortbildungsmaßnahmen über das Phänomen

---

<sup>1</sup> Eine detaillierte Übersicht über die frühen empirischen Befunde zu Reihenfolgeeffekten findet sich bei KLATZKY (1980).

<sup>2</sup> Vgl. LIBBY (1985), S. 654. Für eine ausführliche Beschreibung der durchgeführten analytischen Prüfungshandlung siehe Kap. 3.5.3.3.1.1.2.

<sup>3</sup> Vgl. LIBBY (1985), S. 659.

des recency effects aufzuklären und dessen verzerrenden Einfluss auf die prüferische Urteilsbildung zu verdeutlichen.

### **3.5.3.3.1.2 Einfluss der ersten Hypothese auf die folgende Hypothesengenerierung**

#### **3.5.3.3.1.2.1 Spreading effect**

Psychologische Erkenntnisse weisen darauf hin, dass der Abruf einer bestimmten Hypothese aus dem Gedächtnis die Aktivierung einer mit dieser im engen Zusammenhang stehenden Hypothese wesentlich erleichtert (sog. *spreading effect*).<sup>1</sup> „Stated simply, (...) activating a hypothesis in memory increases the speed of activation of related hypothesis.“<sup>2</sup> „(...) Since similar items are stored together in memory, provision of a cued item from a set of items increases the strength of association in memory for that item, while uncued items are left with relatively weaker associations and are thereby less readily available for recall from memory.“<sup>3</sup>

Im Rahmen der Durchführung von analytischen Prüfungshandlungen würde der spreading effect bedeuten, dass die Aktivierung (Generierung) eines potenziellen Fehlers dazu führen würde, dass anschließend Hypothesen generiert würden, die mit der ersten Hypothese in einem bestimmten Zusammenhang stehen. Die theoretische Grundlage für eine derartige Organisation von Jahresabschlussfehlern bieten Wissensstrukturen an.<sup>4</sup> Diese ordnen Wissensseinheiten (etwa das Wissen über Jahresabschlussfehler) in hierarchischen Ebenen an, wobei die zusammenhängenden Einheiten miteinander verknüpft sind. In Bezug auf die Jahresabschlussprüfung kommt für den Aufbau von Hierarchien insbesondere die Orientierung an Transaktionskreisen (sog. transaction cycles) in Betracht, wie etwa Umsätze & Forderungen oder Einkäufe & Verbindlichkeiten.<sup>5</sup> Folglich würde die Generierung eines bestimmten Fehlers die Aktivierung weiterer Fehler des gleichen Transaktionskreises begünstigen.

---

<sup>1</sup> Vgl. hierzu MEYER/SCHVANEVELDT (1976), S. 27-33.

<sup>2</sup> LIBBY (1985), S. 652. Daher wird synonym auch der Begriff „spreading activation“ verwandt. Etwa bei NG/GREEN/SIMNETT (2001), S. 363.

<sup>3</sup> GREEN (2004a), S. 372.

<sup>4</sup> Vgl. hierzu ausführlich Kap.3.2.2.3.4.2.

<sup>5</sup> Vgl. etwa LIBBY (1985), S. 652; LIBBY/FREDERICK (1990), S. 350.

LIBBY konnte den spreading effect im Rahmen seiner Studie empirisch bestätigen.<sup>1</sup> Prüfer, deren erste Fehlerhypothese dem Transaktionskreis Umsätze & Forderungen (Vorräte & Verbindlichkeiten) zuzuordnen war, generierten anschließend überwiegend Fehlerhypothesen aus dem gleichen Transaktionskreis.<sup>2</sup> Einen vergleichbaren Zusammenhang konnten auch BEDARD/BIGGS feststellen.<sup>3</sup> Sie beobachteten bei einem Teil der Prüfer, dass sie sich im Rahmen ihrer Fehleranalyse verstärkt auf den zuerst in Betracht gezogenen Transaktionskreis konzentrierten und weniger auf andere Transaktionskreise.<sup>4</sup>

Problematisch dabei ist, dass ein Hypothesensatz, welcher sich überwiegend auf einen möglichen Transaktionskreis bezieht, die Gefahr birgt, dass sich die anschließende Informationssuche auch nur auf diesen Aspekt fokussiert.<sup>5</sup> Diese als *Tiefensuche* zu bezeichnende Strategie ist allerdings nur dann vorteilhaft, wenn der tatsächliche Fehler auch diesem Bereich entspringt. Sollte sich der Fehler jedoch in einem anderen Bereich befinden, so könnte sich die Tiefensuche negativ auf die Effektivität und Effizienz der Jahresabschlussprüfung auswirken. Eine wesentlich sinnvollere Vorgehensweise ist in der *Breitensuche* zu sehen, welche sich nicht nur auf einen Transaktionskreis, sondern auf diverse Transaktionskreise konzentriert. Eine derartige Strategie wird jedoch nur dann begünstigt, wenn der Prüfer mehrere Fehlerhypothesen generiert, die sich zudem auf unterschiedliche Bereiche beziehen. Insofern ist der spreading effect für die Prüfungspraxis eher kritisch zu bewerten.

Eine Möglichkeit, diesem Problem entgegenzuwirken, ist etwa darin zu sehen, dass Prüfer explizit dazu aufgefordert werden, diverse auf unterschiedliche Transaktionskreise bezogene Hypothesen zu generieren. Dieses könnte etwa in den Prüfungsstandards oder der Prüfungssoftware der Prüfungsgesellschaften erfolgen. Denkbar wäre auch, dass dem Prüfer in Abhängigkeit der identifizierten unerwarteten Abweichung unterschiedliche potenzielle Fehlerursachen, etwa durch softwaregestützte Entscheidungshilfen, bereitgestellt werden. Nach Möglichkeit sollte der Abschlussprüfer bereits im Rahmen seiner beruflichen Ausbildung, etwa in Mitarbeiterschulungen oder während des Studiums, für die Vorteilhaftigkeit der Breitensuche sensibilisiert werden, so dass er sich regelmäßig um die Generierung von breit angelegten Hypothesensätzen bemüht.

---

<sup>1</sup> Einzelheiten zur Gestaltung der Fallstudie wurden bereits unter Kap. 3.5.3.3.1.2.1 ausführlich diskutiert.

<sup>2</sup> Vgl. LIBBY (1985), S. 654 i.V.m. 662.

<sup>3</sup> Einzelheiten zur Gestaltung der Fallstudie wurden bereits unter Kap. 3.5.2.1 ausführlich diskutiert.

<sup>4</sup> Vgl. BEDARD/BIGGS (1991a), S. 638.

<sup>5</sup> Eine ausführliche Erörterung der Phase der Informationssuche erfolgt unter Kap. 3.5.4.1.

### 3.5.3.3.1.2.2 Interference effect

In einem engen Zusammenhang zum spreading effect steht der sog. *interference effect*. „Interference occurs when an individual is so preoccupied with explanations of one type that he or she cannot adequately consider other types of explanations.”<sup>1</sup> In der psychologischen Forschung wurde der interference effect bereits bei diversen Berufsgruppen nachgewiesen.<sup>2</sup> In Bezug auf analytische Prüfungshandlungen könnte diese „Störung“ dazu führen, dass Prüfer, die zunächst Nicht-Fehlerursachen für eine festgestellte Abweichung generieren, anschließend nicht mehr dazu in der Lage sind, eine adäquate Anzahl von möglichen Fehlerursachen zu generieren.<sup>3</sup> Da die Aufdeckung von Fehlern im Jahresabschluss jedoch das primäre Ziel einer Abschlussprüfung darstellt, könnte der interference effect einen erheblichen Einfluss auf die Effektivität von Jahresabschlussprüfungen mit sich bringen.<sup>4</sup>

Um festzustellen, ob Prüfer bei analytischen Prüfungshandlungen vom interference effect betroffen sind, ließen ANDERSON/KAPLAN/RECKERS 38 Manager mit einer durchschnittlichen Prüfungserfahrung von fünf Jahren eine Fallstudie bearbeiten, bei der Hypothesen für eine wesentliche Abweichung in der Umschlagshäufigkeit (inventory turn over) generiert werden sollten. Dabei wurde eine Gruppe der Prüfer aufgefordert, zunächst ausschließlich potenzielle Nicht-Fehlerursachen (Fehlerursachen) und anschließend Fehlerursachen (Nicht-Fehlerursachen) aufzulisten.<sup>5</sup> Die Ergebnisse von ANDERSON/KAPLAN/RECKERS deuteten auf das Vorhandensein eines asymmetrischen interference effects hin. Prüfer, die zuerst Nicht-Fehlerhypothesen generierten, konnten anschließend signifikant weniger Fehlerhypothesen aktivieren, als Prüfer, die zuerst Fehlerhypothesen generierten. In Bezug auf Nicht-Fehlerhypothesen konnte eine Beeinflussung durch die Reihenfolge allerdings nicht nachgewiesen werden. Prüfer, die zuerst Fehlerhypothesen generierten, waren anschließend in der Lage, eine vergleichbare Anzahl von Nicht-Fehlerhypothesen aufzulisten, wie Prüfer, die zuerst Nicht-Fehlerhypothesen generierten.<sup>6</sup>

---

<sup>1</sup> ANDERSON/KAPLAN/RECKERS (1992), S. 1. Eine vergleichbare Definition findet sich bei NG/GREEN/SIMNETT (2001), S. 356.

<sup>2</sup> Einen Überblick gibt NICKERSON (1984), S. 531-552.

<sup>3</sup> FREDERICK (1991), S. 240-258 konnte den interference effect im Rahmen der Beurteilung des internen Kontrollsystems feststellen.

<sup>4</sup> Ähnlich argumentieren ASHTON/ASHTON (1988), S. 640.

<sup>5</sup> Vgl. ANDERSON/KAPLAN/RECKERS (1992), S. 6-7.

<sup>6</sup> Vgl. ANDERSON/KAPLAN/RECKERS (1992), S. 8.

Für die Prüfungspraxis sollte daher gelten, dass Prüfer bei festgestellten wesentlichen Abweichungen zunächst von Fehlerursachen ausgehen und erst anschließend potenzielle Nicht-Fehlerursachen berücksichtigen sollten. Auf diesem Wege könnte vermieden werden, dass eine potenziell richtige Fehlerursache aufgrund des interference effects nicht im anfänglichen Hypothesensatz enthalten ist und somit die Effektivität der Jahresabschlussprüfung gefährdet wäre. Eine entsprechende Aufforderung durch die Prüfungssoftware der Prüfungsunternehmen wäre eine kostengünstige Maßnahme, durch die einer umgekehrten Reihenfolge entgegengewirkt werden könnte. Entsprechende Hinweise im Rahmen von Schulungsmaßnahmen oder in Lehrbüchern wäre ebenfalls eine Möglichkeit, um Prüfer für die Bedeutung der richtigen Reihenfolge zu sensibilisieren.

### **3.5.3.3.1.3      Anzahl der Hypothesen**

#### **3.5.3.3.1.3.1      Einfluss der Erfahrung auf die Anzahl der generierten plausiblen Hypothesen**

Ein Aspekt, der in der Prüfungsforschung mehrfach untersucht wurde, ist die Anzahl der generierten plausiblen<sup>1</sup> Hypothesen. Dieser Aspekt könnte einen Einfluss auf die Effektivität und Effizienz von analytischen Prüfungshandlungen und somit den gesamten Prüfungsprozess haben. Bei einer geringen Anzahl der plausiblen Hypothesen ist die Wahrscheinlichkeit, dass die korrekte Hypothese berücksichtigt wurde, tendenziell geringer als bei einer hohen Anzahl plausibler Hypothesen. Dem ist entgegenzuhalten, dass eine hohe Anzahl die prüferische Urteilsbildung ggf. unnötig ausweiten und somit die Effizienz der Jahresabschlussprüfung beeinträchtigen könnte. Darüber hinaus können zu viele generierte Hypothesen auch eine Gefährdung der Effektivität verursachen, da sich der Prüfer durch die Vielzahl an zu verarbeitenden Informationen kognitiv überfordert fühlen könnte.<sup>1</sup>

Studien, die sich mit der Anzahl der plausiblen Hypothesen beschäftigten, konzentrierten sich überwiegend auf den Einfluss von Erfahrung. LIBBY/FREDERICK etwa stellten

---

<sup>1</sup> Plausibel bedeutet, dass die Hypothese auf eine Ursache gerichtet ist, die mit der Datenkonstellation im gegebenen Fall übereinstimmt. Dieses bedeutet jedoch nicht immer, dass sie auch der tatsächliche Grund für die festgestellte Abweichung ist, da unterschiedliche Fehlergründe die gleichen Symptome bzw. Muster in den Jahresabschlussdaten verursachen können. Nicht plausible Hypothesen weisen hingegen Inkonsistenzen zu den zugrunde liegenden Datenkonstellationen auf und kommen folglich nicht als tatsächlicher Grund für die festgestellte Abweichung in Betracht.



fest, dass Prüfungsassistenten mit einer durchschnittlichen Prüfungserfahrung von einem Jahr im Schnitt 4,1 plausible Hypothesen generierten, Manager mit einer durchschnittlichen Prüfungserfahrung von fünf Jahren hingegen 5,0.<sup>2</sup> Da erfahrenere Prüfer auch signifikant weniger nicht plausible Hypothesen generierten (0,9 im Vgl. zu 1,5), empfehlen die Autoren, dass analytische Prüfungshandlungen generell von erfahreneren Prüfern durchgeführt werden sollten.<sup>3</sup> ISMAIL/TROTMAN ließen in ihrer Untersuchung ebenfalls 38 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 1,8 Jahren die gleiche Aufgabenstellung bearbeiten.<sup>4</sup> Sie stellten fest, dass die Probanden im Schnitt 5,4 plausible Hypothesen generierten.<sup>5</sup> Im Gegensatz zu den Ergebnissen von LIBBY/FREDERICK generierten die Prüfer in dieser Studie durchschnittlich weniger als 0,3 nicht plausible Hypothesen.<sup>6</sup> BHATTACHARJEE ET AL. konnten im Rahmen einer anderen, jedoch vergleichbaren Studie ähnliche Ergebnisse ermitteln.<sup>7</sup> Die teilnehmenden Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,7 Jahren generierten im Schnitt 5,1 plausible Hypothesen.<sup>8</sup> Tabelle 2 fasst die Ergebnisse zusammen:

| Ø-liche Prüfungserfahrung<br>[Jahren] | Ø-liche Anzahl plausibler<br>Hypothesen | Quelle                      |
|---------------------------------------|-----------------------------------------|-----------------------------|
| 1,0                                   | 4,1                                     | LIBBY/FREDERICK (1990)      |
| 1,8                                   | 5,4                                     | ISMAIL/TROTMAN (1995)       |
| 3,7                                   | 5,1                                     | BHATTACHARJEE ET AL. (1999) |
| 5,0                                   | 5,0                                     | LIBBY/FREDERICK (1990)      |

Tabelle 2: Gegenüberstellung der durchschnittlichen Prüfungserfahrung und der durchschnittlichen Anzahl der generierten plausiblen Hypothesen

Lediglich Prüfer mit durchschnittlich einjähriger Berufserfahrung unterscheiden sich von den Prüfern mit durchschnittlich 1,8, 3,7 bzw. 5,0 Jahren Prüfungserfahrung. Allerdings generieren Prüfer mit einer durchschnittlichen Prüfungserfahrung von 1,8 Jahren im Schnitt mehr als Prüfer mit wesentlich mehr Prüfungserfahrung. Insgesamt lassen die Ergebnisse jedoch keinen eindeutigen Zusammenhang erkennen.<sup>9</sup>

<sup>1</sup> Vgl. SCHREIBER (2000), S. 129-130.

<sup>2</sup> Vgl. LIBBY/FREDERICK (1990), S. 357. Eine ausführliche Darstellung des Versuchsaufbaus wurde bereits unter Kap. 3.5.3.3.1.1.1 vorgenommen.

<sup>3</sup> Vgl. LIBBY/FREDERICK (1990), S. 362.

<sup>4</sup> Vgl. ISMAIL/TROTMAN (1995), S. 349 i.V.m. 351.

<sup>5</sup> Vgl. ISMAIL/TROTMAN (1995), S. 355.

<sup>6</sup> Vgl. ISMAIL/TROTMAN (1995), S. 353.

<sup>7</sup> Hierbei wurde die Fallstudie aus der Studie von BEDARD/BIGGS (1991a) übernommen, welche bereits unter Kap. 3.5.2.1 genauer beschrieben wurde.

<sup>8</sup> BHATTACHARJEE/KIDA/HANNO (1999), S. 90-91.

<sup>9</sup> Allerdings ist einschränkend darauf hinzuweisen, dass die zu bearbeitende Aufgabenstellung bei BHATTACHARJEE/KIDA/HANNO (1999) nicht identisch zu denen der anderen Studien ist. Zwar beschäftigen sich alle herangezogenen Studien mit der Bewertung des Vorratsvermögens, allerdings wurde die festgestellte Abweichung bei BHATTACHARJEE/KIDA/HANNO (1999) durch eine Überbewer-

WRIGHT/WRIGHT nahmen diese konfliktären Ergebnisse zum Anlass, um eine weitere, jedoch modifizierte Studie durchzuführen. Sie gingen davon aus, dass die Fähigkeit zur Generierung von plausiblen Hypothesen nicht entscheidend von der allgemeinen Berufserfahrung abhängt, sondern von der branchenspezifischen Erfahrung. Die Autoren begründen ihre Auffassung damit, dass bestimmte Fehler in Abhängigkeit von der Branche häufiger bzw. seltener auftreten, und dass sich die in diesen Branchen tätigen Prüfer spezialisieren.<sup>1</sup> Daher ließen WRIGHT/WRIGHT 72 Prüfer mit unterschiedlichem Umfang an allgemeiner Prüfungserfahrung und branchenspezifischer Erfahrung an einer Fallstudie teilnehmen, in der es darum ging, eine Prüfungsplanung für ein Handelsunternehmen durchzuführen.<sup>2</sup> Die Fallstudie umfasste insgesamt vier Fehler, von denen drei als branchenspezifisch zu bezeichnen sind.<sup>3</sup> Alle Fehler konnten allein durch die Anwendung von analytischen Prüfungshandlungen aufgedeckt werden.

WRIGHT/WRIGHT fanden heraus, dass Prüfer mit allgemeiner Prüfungserfahrung von durchschnittlich weniger als drei Jahren (bzw. mehr als zwölf Jahren) im Schnitt 4,7 (bzw. 4,2) plausible Hypothesen generierten.<sup>4</sup> Demnach konnte bestätigt werden, dass kein signifikanter Zusammenhang zwischen allgemeiner Berufserfahrung und der Anzahl plausibler Hypothesen vorliegt.<sup>5</sup> Demgegenüber ließen die Ergebnisse einen signifikanten Einfluss von Branchenerfahrung auf die Anzahl der generierten Hypothesen erkennen.<sup>6</sup> Prüfer ohne Branchenerfahrung generierten durchschnittlich 4,11 plausible Hypothesen, wohingegen Prüfer mit umfangreicher Branchenerfahrung (mehr als zwölf branchenspezifische Prüfungsaufträge) durchschnittlich 6,13 plausible Hypothesen generierten.<sup>7</sup> Folglich übt branchenspezifische Erfahrung einen signifikanten Einfluss auf die Anzahl der plausiblen Hypothesen aus.<sup>1</sup>

Berücksichtigt man die Erkenntnisse von WRIGHT/WRIGHT, so scheint eine Spezialisierung nach Branchen durchaus sinnvoll. Daher ist die in der Prüfungspraxis zu beobachtende Branchenspezialisierung von Prüfungsleistungen durchaus positiv zu bewerten.

---

tung der Vorräte verursacht. Bei LIBBY/FREDERICK (1990) sowie ISMAIL/TROTMAN (1995) wurde die festgestellte Abweichung hingegen durch eine Unterbewertung der Vorräte hervorgerufen. Insofern könnte der fehlende Zusammenhang auch auf die nicht identische Aufgabenstellung zurückzuführen sein.

<sup>1</sup> Vgl. WRIGHT/WRIGHT (1997), S. 276.

<sup>2</sup> Vgl. WRIGHT/WRIGHT (1997), S. 278-279.

<sup>3</sup> Vgl. WRIGHT/WRIGHT (1997), S. 279-280.

<sup>4</sup> Vgl. WRIGHT/WRIGHT (1997), S. 285.

<sup>5</sup> Vgl. WRIGHT/WRIGHT (1997), S. 284.

<sup>6</sup> Branchenspezifische Erfahrung wurde dabei anhand der Anzahl der abgewickelten Prüfungsaufträge in einer bestimmten Branche (hier Handelsbranche) gemessen. Vgl. WRIGHT/WRIGHT (1997), S. 284.

<sup>7</sup> Vgl. WRIGHT/WRIGHT (1997), S. 285.

### **3.5.3.3.1.3.2 Einfluss der Anzahl der generierten Hypothesen auf die Effektivität und Effizienz von analytischen Prüfungshandlungen**

Die zuvor herangezogenen Studien untersuchten, inwiefern sich unterschiedliche Erfahrungsarten auf die Anzahl der plausiblen Hypothesen auswirken, ohne dass den Prüfern dabei Vorgaben bezüglich des Umfangs zu generierender Hypothesen gemacht wurden (sog. unconstrained setting). Dabei unterstellten sie einen positiven Zusammenhang zwischen der Anzahl der plausiblen Hypothesen und der Effektivität und Effizienz von analytischen Prüfungshandlungen, indem sie davon ausgingen, dass die Wahrscheinlichkeit des Vorhandenseins der tatsächlichen Ursache bei einer höheren Anzahl von plausiblen Hypothesen größer ist als bei einer geringen Anzahl von plausiblen Hypothesen. Ein empirischer Nachweis für diese Annahme wurde jedoch nicht erbracht.

Um diesen Aspekt genauer zu untersuchen, führten BHATTACHARIE ET AL. eine Studie durch, bei der sie den teilnehmenden Prüfern (insgesamt nahmen 60 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,7 Jahren teil<sup>2</sup>) vorschrieben, wie viele Hypothesen sie anfänglich generieren durften, die ihrer Ansicht nach für eine festgestellte unerwartete Abweichung in Betracht zu ziehen sind.<sup>3</sup> Dabei wurde differenziert zwischen Prüfern, die zur anfänglichen Generierung von einer Hypothese (im Folgenden als Ein-Hypothesen-Prüfer bezeichnet), von drei Hypothesen (im Folgenden als Drei-Hypothesen-Prüfer bezeichnet) und von sechs Hypothesen (im Folgenden als Sechs-Hypothesen-Prüfer bezeichnet) aufgefordert wurden. Einer weiteren Gruppe wurde die Anzahl der anfänglich zu generierenden Hypothesen nicht vorgegeben.<sup>4</sup> Zu betonen ist, dass sich die Vorgabe ausschließlich auf die anfängliche Hypothesengenerierung bezog. Den Prüfern stand es frei, im Rahmen ihrer prüferischen Urteilsbildung weitere Hypothesen zu generieren (sog. anschließende Hypothesengenerierung).

Im Ergebnis konnte festgestellt werden, dass die Ein-Hypothesen-Prüfer im Schnitt insgesamt 2,5 plausible Hypothesen, die Drei-Hypothesen-Prüfer 3,93, die Sechs-Hypothesen-Prüfer 6,53 und die Prüfer ohne Vorgabe 5,07 generierten.<sup>5</sup> Die Effektivität der Drei-Hypothesen-Prüfer war gemeinsam mit den Ein-Hypothesen-Prüfern am

---

<sup>1</sup> Vgl. WRIGHT/WRIGHT (1997), S. 284, 286.

<sup>2</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 87.

<sup>3</sup> Wie bereits erwähnt, wurde hierbei die Fallstudie aus der Studie von BEDARD/BIGGS (1991a) übernommen, welche bereits unter Kap. 3.5.2.1 genauer beschrieben wurde.

<sup>4</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 88.

<sup>5</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 90.

höchsten.<sup>1</sup> Die Drei-Hypothesen-Prüfer benötigten für die Durchführung der Aufgabe zudem signifikant weniger Zeit als die anderen Prüfer (27,81 min. ggü. 36,31 min., 38,07 min. und 38,08 min.).<sup>2</sup> Aufgrund der höchsten Effektivität einerseits und geringsten in Anspruch genommenen Zeit andererseits wiesen die Drei-Hypothesen-Prüfer auch die höchste Effizienz auf. Die Sechs-Hypothesen-Prüfer benötigten unverhältnismäßig viel Zeit für den zuerst generierten Hypothesensatz und wenig Zeit für die anschließend durchgeführten Tests zur Verifizierung bzw. Falsifizierung der Hypothesen, was ihre Effektivität erheblich beeinträchtigte.<sup>3</sup> Vermutlich führte die kognitive Überbeanspruchung (sog. *cognitive overload*) zu einem frühzeitigen Abbruch der Hypothesentests.<sup>4</sup> Dafür spricht auch, dass sie im Schnitt lediglich 1,3 min. Testzeit pro Hypothese (ggü. 4,58 min. bei den Ein-Hypothesen-Prüfern, 1,99 min. bei den Drei-Hypothesen-Prüfern und 2,92 min. bei den Prüfern ohne Beschränkung) in Anspruch nahmen.<sup>5</sup>

Die Ein-Hypothesen-Prüfer benötigten die meiste Zeit für die nachfolgende Hypothesengenerierung und die Hypothesentests. Vermutlich hat die Suche nach und Analyse von Informationen, die zuvor bereits untersucht wurden, zu Doppelarbeiten und somit zu Effizienzverlusten geführt.<sup>6</sup> Andererseits scheint die geringe Anzahl der generierten Hypothesen und die damit verbundene geringe mentale Belastung ein gründlicheres Durchdenken des Problemsachverhalts gewährleistet zu haben, was für ihre hohe Effektivität spricht.<sup>7</sup> Allerdings zeigt die vergleichbare Effektivität bei den Drei-Hypothesen-Prüfern auch, dass der gleiche Zielerreichungsgrad auch mit weniger Zeit erreicht werden konnte. Die Prüfer ohne Vorgaben benötigten im Vergleich zu den Drei-Hypothesen-Prüfern mehr Zeit für die Hypothesentests, ohne dabei die gleiche Effektivität zu erreichen.<sup>8</sup> Diese Ergebnisse nahmen die Autoren zum Anlass, um zu behaupten, dass „evaluating around three initial hypothesis may provide the best balance between the number of multiple hypothesis considered and an auditors’ ability to effectively and efficiently test those hypothesis”.<sup>9</sup>

---

<sup>1</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 98.

<sup>2</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 91.

<sup>3</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 91.

<sup>4</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 98.

<sup>5</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 92.

<sup>6</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 86.

<sup>7</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 97.

<sup>8</sup> Vgl. BHATTACHARJEE/HANNO/KIDA (1999), S. 97.

<sup>9</sup> BHATTACHARJEE/HANNO/KIDA (1999), S. 97.

Für die Prüfungspraxis wäre zu überlegen, Prüfer im Rahmen der Durchführung von analytischen Prüfungshandlungen dazu aufzufordern, lediglich drei für eine festgestellte unerwartete Abweichung in Frage kommende Hypothesen zu generieren oder lediglich die drei wahrscheinlichsten Hypothesen aus dem umfangreicheren Hypothesenbündel einer weiteren Untersuchung zu unterziehen.

#### **3.5.3.3.1.4 Hypothesenwechsel**

Wie in Kap. 3.5.3.3.1.1.1 dargestellt, neigen Prüfer dazu, häufig auftretende Fehler (frequently occurring errors) als Grund für eine festgestellte Abweichung zu generieren. Eine derartige Vorgehensweise kann zu einer effektiven und effizienten Prüfung beitragen, wenn es sich bei der anfänglichen Hypothese um den tatsächlichen Grund handelt. Sollte dieses jedoch nicht der Fall sein (etwa weil es sich um einen weniger häufig auftretenden Fehler handelt), so hängt der weitere Erfolg des Prüfers maßgeblich davon ab, ob er die Unzweckmäßigkeit seiner Hypothesen erkennt und diese durch eine effektivere Hypothesengenerierung ersetzt.<sup>1</sup> Die Ergebnisse der empirischen Prüfungsforschung zeigen allerdings, dass Prüfer Schwierigkeiten haben, eine anfänglich generierte Hypothese zu verwerfen und durch eine andere zu ersetzen.<sup>2</sup> BEDARD/BIGGS etwa fanden heraus, dass die mangelnde Fähigkeit von Prüfern, eine anfängliche Hypothese zu wechseln, einen Hauptgrund für die schlechte Erfolgsquote bei der Durchführung einer analytischen Prüfungshandlung darstellt.<sup>3</sup> Die teilnehmenden Prüfer waren nicht einmal dann zu einem Hypothesenwechsel in der Lage, wenn gegenteilige Beweise erlangt wurden.<sup>4</sup> Vergleichbare Ergebnisse wurden auch von ASARE/WRIGHT festgestellt.<sup>5</sup>

Um zu eruieren, von welchen Faktoren der Hypothesenwechsel von Prüfern maßgeblich beeinflusst wird, führten HEIMAN-HOFFMAN/MOSER/JOSEPH eine Untersuchung mit 54 Prüfern mit einer durchschnittlichen Prüfungserfahrung von 3 Jahren durch.<sup>1</sup> Die Prüfer wurden aufgefordert, eine Hypothese für eine unerwartete Abweichung zu generieren, welche durch eine Überwertung der Umsatzerlöse i.H.v. 1,5 Mio. \$ verursacht wurde. Anschließend wurden 50% der Prüfer Informationen zur Verfügung gestellt, die auf ei-

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 131.

<sup>2</sup> An dieser Stelle soll auch daran erinnert werden, dass eine Generierung von weiteren Hypothesen möglicherweise durch den interference effect erheblich beeinträchtigt wird. Siehe ausführlich Kap. 3.5.3.3.1.2.2.

<sup>3</sup> Vgl. BEDARD/BIGGS (1991a), S. 638-639.

<sup>4</sup> Vgl. BEDARD/BIGGS (1991a), S. 640.

<sup>5</sup> Vgl. ASARE/WRIGHT (2003), S. 235.

nen häufigen Fehler in den Umsätzen (Abgrenzungsfehler) hindeuteten, und die anderen 50% bekamen hingegen Informationen, die einen seltenen Fehler in den Umsätzen (Doppelbuchung) erkennen ließen. Daraufhin wurden die Prüfer erneut aufgefordert, eine Hypothese für die Abweichung zu generieren, wobei es ihnen erlaubt war, an ihrer ersten Hypothese festzuhalten oder eine andere Hypothese zu benennen.<sup>2</sup>

Als erstes konnten die Autoren feststellen, dass die Probanden überwiegend den Abgrenzungsfehler generierten. Dieser wurde insgesamt 32-mal, also von knapp 60% der Prüfer, als Fehlerursache gewählt. Die anderen Prüfer generierten hingegen diverse andere Fehler, von denen keiner häufiger als fünfmal generiert wurde.<sup>3</sup> Insofern konnten die bisherigen Ergebnisse von LIBBY sowie LIBBY/FREDERICK erneut bestätigt werden.<sup>4</sup>

Weiterhin ging aus den Ergebnissen hervor, dass die Effektivität der Prüfer am höchsten war, wenn sie einen häufigen Fehler generierten und der tatsächliche Grund ebenfalls ein häufiger Fehler war (Erfolgsquote=100%). Allerdings war die Erfolgsquote am schlechtesten, wenn die Prüfer einen häufigen Fehler generierten, es sich bei dem tatsächlichen Grund allerdings um einen seltenen Fehler handelte (Erfolgsquote=22%). Ein weitaus besseres Ergebnis erreichten hingegen die Prüfer, die einen seltenen Fehler generierten, wenn es sich bei dem tatsächlichen Grund um einen häufigen Fehler (Erfolgsquote=83%) oder um einen anderen seltenen Fehler (Erfolgsquote=40%) handelte.<sup>5</sup> Demnach waren die Schwankungen in der Erfolgsquote bei einem (erst-)generierten häufigen Fehler wesentlich größer (100% ggü. 22%) als bei einem (erst-)generierten seltenen Fehler (83% ggü. 40%).

Den maßgeblichen Grund für diese Unterschiede sehen die Autoren in der Fähigkeit bzw. Bereitschaft der Prüfer, von einem seltenen Fehler zu einem häufigen Fehler oder zu einem anderen seltenen Fehler zu wechseln. Den Prüfern bereitete es jedoch überwiegend Probleme, einen häufigen Fehler gegen einen seltenen Fehler zu ersetzen, auch dann, wenn gegenteilige Informationen vorlagen.<sup>6</sup> Für die Prüfungspraxis könnte dieser Aspekt erhebliche Beeinträchtigungen der Effektivität und Effizienz bedeuten. Sollte der Prüfer eine Hypothese generieren, die auf einen häufig vorkommenden Fehler weist, obwohl es sich um einen anderen, selten vorkommenden Fehler handelt, so würde er

---

<sup>1</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 767.

<sup>2</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 770-771.

<sup>3</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 771.

<sup>4</sup> Vgl. hierzu Kap. 3.5.3.3.1.1.1.

<sup>5</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 772.

<sup>6</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 775.

diesen möglicherweise nicht entdecken oder er würde durch weitere, möglicherweise sogar kostenintensive Prüfungshandlungen feststellen, dass seine Vermutung keinen Bestand hat. Hätte ihn die gegenteilige Beweislage jedoch schon vorher zu einem Hypothesenwechsel veranlasst, so wären ihm diese zusätzlichen Kosten erspart geblieben und Effizienzverluste wären nicht entstanden.

Eine Möglichkeit, diesen Gefährdungen entgegenzuwirken, ist darin zu sehen, dass sich Prüfer über diese potenziellen Verzerrungen bewusst werden. Dieses könnte dadurch erreicht werden, dass man ihnen im Rahmen von Schulungsmaßnahmen oder durch Fachzeitschriften bzw. Lehrbücher die Notwendigkeit des iterativen prüferischen Vorgehens verdeutlicht, bei dem u.U. erstgenerierte Hypothesen verworfen und durch neue Hypothesen ersetzt werden müssen.<sup>1</sup>

### **3.5.3.3.2 Externe Hypothesengenerierung**

#### **3.5.3.3.2.1 Der Einfluss des Erhalts einer Hypothese durch den Mandanten bzw. dessen Mitarbeiter**

##### **3.5.3.3.2.1.1 Vorbemerkungen**

Eines der kontroversesten Themen in der Prüfungsforschung ist die Zweckmäßigkeit der Befragung des Mandanten nach potenziellen Gründen für eine festgestellte Abweichung.<sup>2</sup> Diverse Studien aus den frühen Jahren der Prüfungsforschung wiesen darauf hin, dass der Erhalt einer sog. Mandantenerklärung die weitere interne Generierung von Hypothesen maßgeblich beeinflussen und somit einen negativen Einfluss auf die Effektivität und Effizienz von Jahresabschlussprüfungen ausüben könnte.<sup>3</sup> KINNEY/HAYNES waren die Ersten, die dieses zum Anlass nahmen, um auf einen möglichen Missstand bei der Durchführung von analytischen Prüfungshandlungen in der Prüfungspraxis hinzuweisen.<sup>4</sup> Sie befürchteten, dass Mandanten bei Nachfrage durch den Prüfer tendenziell dazu neigen, Nicht-Fehlerhypothesen für den Grund einer festgestellten Abweichung zu liefern.<sup>5</sup> Da eine derartige Vorgehensweise damals wie auch heute von den

---

<sup>1</sup> So auch ASARE/WRIGHT (2003), S. 236.

<sup>2</sup> So auch KOONCE (1993), S.63.

<sup>3</sup> Siehe auch Kap. 3.5.2.2.1.

<sup>4</sup> Vgl. KINNEY/HAYNES (1990), S. 95-101.

<sup>5</sup> Vgl. KINNEY/HAYNES (1990), S. 96. Die gleiche Befürchtung äußerten auch ANDERSON/KAPLAN/RECKERS (1992), S. 10.

einschlägigen Prüfungsstandards empfohlen wird<sup>1</sup> und somit zur gängigen Praxis gehört<sup>2</sup>, schlugen die Autoren vor, die Prüfungsstandards entsprechend zu modifizieren.<sup>3</sup> Diese Aufforderungen führten zu einer Vielzahl von Studien, die sich mit dem Einfluss einer inkorrekten Nicht-Fehlererklärung des Mandanten auf die eigene Hypothesenbildung beschäftigten.

Als (verhaltens-)theoretische Grundlage möglicher Verzerrungen untersuchten diese Studien insbesondere den interference effect und den spreading effect, die bereits im Rahmen der internen Hypothesengenerierung ausführlich erläutert wurden.<sup>4</sup> Bezogen auf den Erhalt einer Mandantenerklärung könnten diese beiden Phänomene dazu führen, dass der Prüfer bei der nachfolgenden Hypothesengenerierung überwiegend Nicht-Fehlerursachen abrufte (spreading effect) und des Weiteren weniger oder keine Fehlerhypothesen generieren kann (interference effect).

### **3.5.3.3.2.1.2 Einfluss auf die nachfolgende interne Hypothesengenerierung**

#### **3.5.3.3.2.1.2.1 Interference effect**

Eine der ersten Studien, die im Zusammenhang mit der Beeinflussung durch Mandantenerklärungen zu sehen ist, stellt die Untersuchung von HEIMAN dar.<sup>5</sup> Im Ergebnis stellte die Autorin einen signifikant negativen Einfluss einer „vererbten“ Hypothese auf die anschließende eigene Hypothesenbildung fest.<sup>6</sup> Obwohl es sich hierbei um eine regelmäßig zitierte Studie zum interference effect beim Erhalt einer Mandantenerklärung handelt, lässt sich durch eine genaue Betrachtung des Untersuchungsdesigns jedoch eindeutig feststellen, dass es sich hierbei nicht explizit um eine Mandantenerklärung handelte.<sup>7</sup> Trotz dieser methodischen Unzulänglichkeit weist HEIMAN auf die potenzielle Gefahr eines durch eine Nicht-Fehlererklärung des Mandanten ausgelösten interference effects hin.<sup>8, 1</sup>

---

<sup>1</sup> Vgl. IDW PS 312.27, ISA 520.18.

<sup>2</sup> Vgl. auch HIRST/KOONCE (1996), S. 462, 471.

<sup>3</sup> Vgl. KINNEY/HAYNES (1990), S. 99.

<sup>4</sup> Vgl. hierzu ausführlich Kap. 3.5.3.3.1.2.2 sowie 3.5.3.3.1.2.1.

<sup>5</sup> Vgl. HEIMAN (1990), S. 875-890.

<sup>6</sup> Vgl. HEIMAN (1990), S. 880 i.V.m. 883.

<sup>7</sup> Siehe HEIMAN (1990), S. 881: „All subjects were asked the question, ‘Given the fact that a financial statement error occurred and effected the ratios on the previous page, how likely do you think it is that the error was *next period’s credit sales being recorded in the current period?*’”

<sup>8</sup> Vgl. HEIMAN (1990), S. 889.



Eine Studie, welche die teilnehmenden Prüfer explizit erkennen ließ, dass es sich um eine Mandantenerklärung handelte, führten NG/GREEN/SIMNETT durch. Sie ließen insgesamt 91 Prüfer mit einer Prüfungserfahrung von 1,25 bis sieben Jahren (durchschnittlich 3,1 Jahre) an ihrer Untersuchung teilnehmen.<sup>2</sup> Die zu bearbeitende Aufgabenstellung umfasste einen nicht erfassten Beschaffungsvorgang im Bereich der Vorräte, was zu einer Unterbewertung der Verbindlichkeiten i.H.v. 12,5 Mio. \$ (5% der Bilanzsumme) führte. Die Aufgabe der Probanden bestand darin, potenzielle Fehlerhypothesen für die verursachten Abweichungen im Jahresabschluss zu generieren.<sup>3</sup> Einer Gruppe von Prüfern wurde dabei vor der eigenen Hypothesengenerierung eine Nicht-Fehlererklärung des Mandanten mitgeteilt, einer weiteren Prüfergruppe hingegen nicht.<sup>4</sup> Die Ergebnisse wiesen darauf hin, dass kein interference effect nachgewiesen werden konnte. Prüfer mit Mandantenerklärung generierten vergleichbar viele Fehlerhypothesen ( $\bar{X} = 4,02$ ) wie Prüfer ohne Mandantenerklärung ( $\bar{X} = 4,11$ ).<sup>5</sup>

Eine weitere Studie, die eine explizite Betrachtung des interference effects zulässt, führte GREEN durch. Sie ließ insgesamt 63 Prüfer mit einer Prüfungserfahrung von drei bis sechs Jahren ( $\bar{X} = 3,84$  Jahre) einen Fall bearbeiten, bei dem die gesamte Plankostenabweichung zum Ende des Geschäftsjahrs den Umsatzkosten zugerechnet wurde, anstatt diese zwischen Umsatzkosten und dem Vorratsvermögen zu verteilen. Dieses führte zu einer Überbewertung der Umsatzkosten und Unterbewertung des Vorratsvermögens sowie des Jahresüberschusses.<sup>6</sup> Die teilnehmenden Prüfer wurden aufgefordert, eine potenzielle Hypothese für die daraus resultierende Abweichung in der Bruttogewinnmarge zu generieren. Die teilnehmenden Prüfer wurden zufällig in drei Gruppen aufgeteilt. Die erste Prüfergruppe erhielt eine Nicht-Fehlererklärung des Mandanten bevor sie zur Hypothesengenerierung aufgefordert wurde. Der zweiten Gruppe wurde die gleiche Mandantenerklärung nach der eigenen Hypothesengenerierung mitgeteilt, wohingegen der dritten Gruppe keine Erklärung zur Verfügung gestellt wurde.<sup>7</sup> Im Ergebnis konnten die Autoren keine signifikanten Unterschiede in der Anzahl der selbstgenerierten Fehlerhypothesen feststellen (2,43 ggü. 2,67 bzw. 2,14), was darauf schließen lässt, dass die

---

<sup>1</sup> Auch an dieser Stelle soll nochmals explizit darauf hingewiesen werden, dass sich die aufgeführte Kritik ausschließlich auf die Übertragung der Ergebnisse auf eine Mandantenerklärung und nicht auf das Untersuchungsdesign an sich bezieht.

<sup>2</sup> Vgl. NG/GREEN/SIMNETT (2001), S. 358-359.

<sup>3</sup> Vgl. NG/GREEN/SIMNETT (2001), S. 357.

<sup>4</sup> Vgl. NG/GREEN/SIMNETT (2001), S. 358.

<sup>5</sup> Vgl. NG/GREEN/SIMNETT (2001), S. 361.

<sup>6</sup> Vgl. GREEN (2004a), S. 375-376.

<sup>7</sup> Vgl. GREEN (2004a), S. 375.

Prüfer, die eine Mandantenerklärung vor der eigenen Hypothesengenerierung erhielten, auch hier nicht vom interference effect betroffen waren.<sup>1</sup>

#### **3.5.3.3.2.1.2.2 spreading effect**

Neben der Untersuchung des interference effects ging GREEN in ihrer Studie auch der Frage nach, ob sich der Erhalt einer Nicht-Fehlererklärung durch den Mandanten möglicherweise dahingehend auswirkt, dass die nachfolgende interne Hypothesengenerierung durch einen verstärkten Abruf weiterer Nicht-Fehlerursachen des gleichen Transaktionskreises geprägt ist.<sup>2</sup> Wie in Kap. 3.5.3.3.1.2.1 gezeigt wurde, konnte dieser sog. spreading effect bereits im Rahmen der internen Hypothesengenerierung nachgewiesen werden. GREEN konnte in ihrer Studie zwar nicht feststellen, dass es sich bei den nach Erhalt der Mandantenerklärung selbstgenerierten Hypothesen überwiegend um Hypothesen handelt, die sich auf den gleichen Transaktionskreis beziehen. Dennoch wiesen die Ergebnisse auf eine wesentliche Beeinflussung durch die Mandantenerklärung hin. Prüfer, denen eine Mandantenerklärung erteilt wurde, generierten signifikant mehr Nicht-Fehlerhypothesen ( $\bar{X} = 4,0$ ) als Prüfer ohne Mandantenerklärung ( $\bar{X} = 3,1$ ).

#### **3.5.3.3.2.1.2.3 Bewertung der Forschungsergebnisse zum interference effect und spreading effect**

Möchte man die Forschungsergebnisse zum Einfluss einer Nicht-Fehlererklärung des Mandanten auf die nachfolgende interne Hypothesengenerierung bewerten, so sind zwei wesentliche Aspekte zu berücksichtigen. Zum einen ist zu konstatieren, dass der im Rahmen der internen Hypothesenbildung festgestellte interference effect lediglich in der Studie von HEIMAN, deren Validität aufgrund des Untersuchungsdesigns fraglich ist, nachgewiesen werden konnte. Nicht evident war der interference effect indes in den Studien von NG/GREEN/SIMNETT und GREEN. Aus diesem Grund kann der häufig geäußerten Befürchtung, dass die von den Prüfungsstandards nahe gelegte Befragung des Mandanten zu einer eingeschränkten Generierung von Fehlerhypothesen führe, nur bedingt zugestimmt werden. Zum anderen ist festzuhalten, dass sich eine Nicht-Fehlererklärung des Mandanten dahingehend auf die nachfolgende interne Hypothesen-

---

<sup>1</sup> Vgl. GREEN (2004a), S. 379-380.

<sup>2</sup> Vgl. GREEN (2004a), S. 373. Für eine detaillierte Darstellung der Untersuchung sei auf die entsprechenden Ausführungen unter Kap. 3.5.3.3.2.1.2.1 verwiesen.

generierung auswirkt, dass es zu einem verstärkten Abruf weiterer Nicht-Fehlerhypothesen kommt (spreading effect). Insofern kann auf Grundlage der bisherigen Befunde keine abschließende Aussage über die Zweckmäßigkeit der Einholung einer Mandantenerklärung getroffen werden.

### **3.5.3.3.2.1.3 Der Einfluss des Erhalts einer Mandantenerklärung auf den Erfolg einer analytischen Prüfungshandlung**

#### **3.5.3.3.2.1.3.1 Vorbemerkungen**

Um die vollständige Tragweite des Erhalts einer Mandantenerklärung einschätzen zu können, reicht es nicht aus, lediglich den Einfluss auf die nachfolgende interne Hypothesengenerierung zu betrachten. Vielmehr gilt es zu untersuchen, inwiefern sich der Erhalt einer Mandantenerklärung auf den Gesamterfolg der analytischen Prüfungshandlung auswirkt. Studien, die genau diesen Aspekt durchleuchten, können dabei in zwei Klassen unterteilt werden. Zum einen handelt es sich dabei um Studien, die keine Kontrollgruppe berücksichtigen, d.h. alle teilnehmenden Prüfer erhalten eine Mandantenerklärung. Zum anderen handelt es sich um Studien, deren Untersuchungsaufbau explizit eine Kontrollgruppe integrieren. D.h., es werden mindestens zwei Prüfergruppen gebildet, die sich darin unterscheiden, ob eine Mandantenerklärung zur Verfügung gestellt wird oder nicht.

#### **3.5.3.3.2.1.3.2 Untersuchungsdesign ohne Kontrollgruppe**

Eine Untersuchung, die keine Kontrollgruppe berücksichtigte, führten BIERSTAKER/BEDARD/BIGGS durch.<sup>1</sup> Sie ließen allen zwölf teilnehmenden Prüfern eine Nicht-Fehlererklärung des Mandanten zukommen. Anschließend war lediglich einer der zwölf Prüfer selbstständig dazu in der Lage, die Nicht-Korrektheit der Mandantenerklärung zu erkennen und die analytische Prüfungshandlung erfolgreich abzuschließen. Allen übrigen Prüfern war es hingegen nicht möglich, die falsche Mandantenerklärung eigenständig zu verwerfen.

BIERSTAKER/BEDARD/BIGGS beobachteten weiterhin, dass das weitere Vorgehen der nicht erfolgreichen Prüfer durch zwei wesentliche Verhaltensmuster gekennzeichnet

---

<sup>1</sup> Für eine genaue Beschreibung der Fallstudie und der Aufgabenstellung siehe Kap. 3.5.2.3 i.V.m. 3.5.2.1.

war. Erstens evaluierten sie die Mandantenerklärung immer und immer wieder (sog. wheel spinning), obwohl bereits Informationen vorlagen, die auf die Falschheit der Mandantenerklärung hindeuteten. Zweitens war es ihnen nicht möglich, eine neue Abweichungshypothese zu generieren, obwohl die Mandantenerklärung aufgrund gegenteiliger Informationen bereits verworfen wurde (sog. truncated hypothesis generation).<sup>1</sup>

Diese beiden „Störungen“ weisen eindeutig auf einen wesentlichen Einfluss einer Mandantenerklärung hin, bei denen im Übrigen – und hier insbesondere dem Phänomen der truncated hypothesis generation – eine gewisse Parallele zum interference effect nicht abgestritten werden kann. Zwar ist die Studie von BIERSTAKER/BEDARD/BIGGS aufgrund des unterschiedlichen Untersuchungsdesigns nicht direkt mit den Ergebnissen von NG/GREEN/SIMNETT und GREEN – welche keinen interference effect nachweisen konnten – zu vergleichen. Dennoch ist den Ergebnissen von BIERSTAKER/BEDARD/BIGGS eine höhere Bedeutung beizumessen, da sie sich nicht ausschließlich auf die Beobachtung der Hypothesengenerierung beschränken, sondern vielmehr den Erfolg und mithin den gesamten Prozess der analytischen Prüfungshandlung betrachten. Aufgrund dieser ganzheitlichen Betrachtungsweise ist es ihnen zugleich möglich, auf eine der Hypothesengenerierung nachgelagerten Phase einzugehen. Das von den Autoren identifizierte wheel spinning weist unverkennbar darauf hin, dass sich der Erhalt einer Mandantenerklärung verzerrend auf die Informationssuche auswirken kann. Eine genauere Betrachtung des Einflusses einer Mandantenerklärung auf die Informationssuche soll hier jedoch nicht vorgenommen werden und wird an entsprechender Stelle nachgeholt.<sup>2</sup>

### **3.5.3.3.2.1.3.3 Untersuchung mit Kontrollgruppe**

Ein wesentlicher Kritikpunkt an der zuvor dargestellten Studie ohne Kontrollgruppe ist in der eindimensionalen Betrachtungsweise zu sehen. Es wird lediglich die Vorgehensweise bzw. der Erfolg von Probanden untersucht, denen eine inkorrekte Mandantenerklärung mitgeteilt wurde. Auf diesem Wege ist es zwar möglich, das Zustandekommen eines fehlerhaften Urteils zu analysieren. Doch obwohl es sich hierbei um ein höchst wichtiges Erkenntnisinteresse handelt, bleibt bei einem derartigen Untersuchungsdesign unbeantwortet, ob Prüfer die gleiche Aufgabe erfolgreicher bearbeitet hätten, wenn ihnen keine inkorrekte Mandantenerklärung mitgeteilt worden wäre. Genau diese

---

<sup>1</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 32-33.

<sup>2</sup> Siehe hierzu ausführlich Kap. 3.5.4.4.3.

Schwachstellen beheben Untersuchungen mit mindestens einer Kontrollgruppe. Der Vorteil dieses differenzierteren Untersuchungsdesigns ist in erster Linie darin zu sehen, dass die Vorgehensweise bzw. der Erfolg der unterschiedlichen Gruppen miteinander verglichen werden kann.

Eine der wenigen Studien, die nicht ausschließlich den Erfolg von Prüfern untersuchte, denen eine Mandantenerklärung mitgeteilt wurde, führten BEDARD/BIGGS durch.<sup>1</sup> Die Autoren werteten insgesamt 176 bearbeitete Fallstudien aus. Davon wurden 97 von Prüfungsleitern mit einer durchschnittlichen Berufserfahrung von 3,5 Jahren und 79 von Managern mit einer durchschnittlichen Berufserfahrung von 7,2 Jahren bearbeitet.<sup>2</sup> Es wurden insgesamt drei Prüfergruppen gebildet, die sowohl Prüfungsleiter als auch Manager umfassten. 55 Prüfer erhielten einen Fall ohne Mandantenerklärung, 62 Prüfer einen Fall mit einer inkorrekten Nicht-Fehlererklärung des Mandanten und 59 Prüfer einen Fall mit einer korrekten Fehlererklärung des Mandanten.<sup>3</sup>

BEDARD/BIGGS stellten entgegen ihren Erwartungen fest, dass die Nicht-Fehlererklärung des Mandanten keinen signifikant negativen Einfluss auf den Erfolg der Prüfer hatte. Die Prüfergruppe, die eine inkorrekte Nicht-Fehlererklärung des Mandanten erhielt, wies eine vergleichbare Erfolgsquote zu den Prüfern auf, die keine Mandantenerklärung zur Verfügung bekommen haben (32% ggü. 35%).<sup>4</sup> Die Autoren nahmen ihre Ergebnisse zum Anlass, den in der Prüfungsforschung häufig als effektivitäts- und effizienzmindernden interference effect beim Erhalt einer Mandantenerklärung in Frage zu stellen.<sup>5</sup>

Allerdings bleibt einschränkend festzuhalten, dass diese Studie aufgrund der zur Verfügung stehenden Daten keine explizite Aussage bezüglich der (Nicht-)Existenz des interference effects erlaubt.<sup>6</sup> Es wurde weder die Anzahl noch die Art der selbstgenerierten Hypothesen erfasst. Von den Probanden wurde lediglich die Beschreibung einer einzi-

---

<sup>1</sup> Vgl. BEDARD/BIGGS (1991b), S. 81-82.

<sup>2</sup> Vgl. BEDARD/BIGGS (1991b), S.82. Da es sich bei der weiteren Ausgestaltung der Aufgabenstellung um den gleichen Aufbau wie bei der Studie von BEDARD/BIGGS (1991a) handelt, sei an dieser Stelle auf die entsprechenden Ausführungen unter Kap. 3.5.2.1 verwiesen.

<sup>3</sup> Vgl. BEDARD/BIGGS (1991b), S. 83.

<sup>4</sup> Vgl. BEDARD/BIGGS (1991b), S. 83. Auf eine Analyse der Prüfer, die eine richtige Mandantenerklärung erhielten, soll an einer späteren Stelle dieses Kapitels eingegangen werden.

<sup>5</sup> Vgl. BEDARD/BIGGS (1991b), S. 88.

<sup>6</sup> Die Fallstudie wurde den Teilnehmern über eine Kontaktperson der entsprechenden Wirtschaftsprüfungsunternehmen übergeben. Die Probanden wurden aufgefordert, den Fall alleine zu bearbeiten. Sie wurden weder beobachtet, noch dazu aufgefordert, think aloud-Protokolle zu führen. Anschließend wurde der bearbeitete Fall den Autoren zugesandt. Eine Datenanalyse konnte also ausschließlich anhand der vorliegenden Fallunterlagen erfolgen. Vgl. BEDARD/BIGGS (1991b), S. 82.

gen und zwar der wahrscheinlichsten Abweichungsursache verlangt. Dezidierte Daten zur Untersuchung eines interference effects lagen hingegen nicht vor.

#### **3.5.3.3.2.1.3.4 Bewertung**

Auch die Analyse der Studien von BIERSTAKER/BEDARD/BIGGS auf der einen Seite und von BEDARD/BIGGS auf der anderen Seite lässt (noch) keine eindeutige Aussage über die Zweckmäßigkeit des Erhalts einer Mandantenerklärung zu. Die erste Autorengruppe stellte eindeutig fest, dass die Mitteilung einer falschen Mandantenerklärung erhebliche „Störungen“ der weiteren prüferischen Vorgehensweise mit sich bringt und lediglich einer von zwölf teilnehmenden Prüfern die analytische Prüfungshandlung eigenständig erfolgreich abschließen konnte. Auch die zweite Autorengruppe stellte fest, dass lediglich 32% der Prüfer mit inkorrektur Mandantenerklärung in der Lage waren, die analytische Prüfungshandlung erfolgreich abzuschließen. Allerdings konnten sie auch nachweisen, dass Prüfer ohne Mandantenerklärung insgesamt nicht signifikant erfolgreicher waren (35%). Aus diesem Blickwinkel würde ein Plädoyer für die Änderung der durch IDW PS 312 sowie ISA 520 vorgeschlagenen prüferischen Vorgehensweise bei der Feststellung einer wesentlichen unerwarteten Abweichung kaum gerechtfertigt sein.

#### **3.5.3.3.2.1.4 Der Einfluss der Prüfererfahrung auf den Erfolg analytischer Prüfungshandlungen bei Erhalt einer Mandantenerklärung**

##### **3.5.3.3.2.1.4.1 Einfluss von allgemeiner und domänenspezifischer Erfahrung**

Um eine weitergehende Klärung der Zweckmäßigkeit einer Mandantenbefragung bemüht, erweiterten BEDARD/BIGGS ihr Untersuchungsdesign in zwei wesentlichen Punkten. Zum einen vermuteten sie, dass der Einfluss einer inkorrektur Mandantenerklärung auf den Erfolg eines Prüfers in Abhängigkeit der jeweiligen Prüfererfahrung zu betrachten sei. Um diesen Aspekt genauer zu durchleuchten, klassifizierten BEDARD/BIGGS die teilnehmenden Prüfer einerseits nach der allgemeinen Prüfungserfahrung (als Kriterium wurde hierbei die Anzahl der Berufsjahre im Prüfungswesen herangezogen) und andererseits nach der dömanenspezifischen Erfahrung.<sup>1</sup> Da es sich bei dem Grund für die festgestellte Abweichung um einen Fehler im Bereich des Vorratsvermögens handelte,

---

<sup>1</sup> Eine ausführliche Beschreibung der Studie von BEDARD/BIGGS (1991b) ist dem Kap. 3.5.3.3.2.1.3.3 zu entnehmen.

erfolgte eine Einteilung nach domänenspezifischer Erfahrung anhand der Spezialisierung der Prüfer auf Mandanten, die im Allgemeinen erhebliche Vorratsbestände aufweisen (etwa Produktions- oder Handelsunternehmen).<sup>1</sup> Zum anderen teilten BEDARD/BIGGS die Probanden nicht nur in eine Gruppe mit und ohne inkorrekt Mandantenerklärung ein. Vielmehr bildeten die Autoren eine dritte Prüfergruppe, denen eine Mandantenerklärung zugeteilt wurde, welche die wahre Fehlerursache umfasste.

Zunächst einmal wiesen BEDARD/BIGGS nach, dass der Grad der allgemeinen Prüfungserfahrung keinen wesentlichen Einfluss auf den Erfolg der analytischen Prüfungshandlungen hatte.<sup>2</sup> Bei den Prüfern mit einer durchschnittlichen Berufserfahrung von (1) weniger als drei Jahren, (2) zwischen drei und sechs Jahren und (3) mehr als sechs Jahren konnten keine signifikanten Erfolgsunterschiede nachgewiesen werden (42% bei (1) ggü. 48% bei (2) und 49% bei (3)).<sup>3</sup> Klassifiziert man die Prüfer hingegen nach der domänenspezifischen Erfahrung, so zeichnet sich ein erheblich anderes Bild ab. Von den Prüfern mit einem hohen Anteil an vorratsintensiven Mandanten bearbeiteten 57% die Aufgabe erfolgreich, wohingegen lediglich 40% der Prüfer mit weniger vorratsintensiven Mandanten die Aufgabe richtig lösten.<sup>4</sup>

Um den Grund für ihre Ergebnisse genauer zu analysieren, differenzierten BEDARD/BIGGS zusätzlich nach der jeweils zugrunde liegenden Aufgabenstellung.<sup>5</sup> Dabei erkannten die Autoren, dass „domain-specific experience was most associated with recognizing the fit of an ‘inherited’ hypothesis to a pattern of discrepancies (...)“.<sup>6</sup> Bei Erhalt einer korrekten Mandantenerklärung lösten 92% der Prüfer mit höherer Domänenenerfahrung und lediglich 57% der Prüfer mit weniger Domänenenerfahrung die Aufgabe richtig. Die Unterschiede zwischen domänenenerfahrenen und -unerfahrenen Prüfern waren hingegen nicht signifikant, wenn die Prüfer eine inkorrekte Mandantenerklärung (36% ggü. 30%) oder keine Mandantenerklärung (36% ggü. 33%) erhielten.<sup>7</sup>

Für die Prüfungspraxis lassen sich die Ergebnisse von BEDARD/BIGGS in zweifacher Hinsicht deuten. Erstens, unabhängig von der Domänenspezifität führt der Erhalt einer

---

<sup>1</sup> Vgl. BEDARD/BIGGS (1991b), S. 82.

<sup>2</sup> Vgl. BEDARD/BIGGS (1991b), S. 85.

<sup>3</sup> Vgl. BEDARD/BIGGS (1991b), S. 86. Vergleichbare Ergebnisse ergaben sich bei einer Klassifizierung nach Prüfungsleitern (41%) und Managern (53%).

<sup>4</sup> Vgl. BEDARD/BIGGS (1991b), S. 86.

<sup>5</sup> Wie bereits erwähnt, wurde einem Drittel der Prüfer eine inkorrekte Nicht-Fehlererklärung des Mandanten mitgeteilt, einem weiteren Drittel hingegen eine korrekte Fehlererklärung. Dem letzten Drittel der Prüfer wurde keine Mandantenerklärung zur Verfügung gestellt.

<sup>6</sup> BEDARD/BIGGS (1991b), S. 86.

<sup>7</sup> Vgl. BEDARD/BIGGS (1991b), S. 87.

Mandantenerklärung im Vergleich zur „reinen“ internen Hypothesengenerierung nicht zu einer schlechteren Erfolgsquote. Prüfer mit einer falschen Mandantenerklärung sind ähnlich erfolgreich wie Prüfer ohne Mandantenerklärung (36% & 30% ggü. 36% & 33%). War die Mandantenerklärung hingegen korrekt, so waren die Prüfer sogar wesentlich erfolgreicher als Prüfer ohne Mandantenerklärung (92% & 57% ggü. 36% & 33%). Insofern sprechen mehr Argumente für als gegen eine Mandantenbefragung.

Zweitens, nimmt man an, dass die in der Prüfungspraxis zu beobachtende Branchenspezialisierung der Prüfungsgesellschaften dazu führt, dass analytische Prüfungshandlungen überwiegend von domänen erfahrenen Prüfern durchgeführt werden<sup>1</sup>, so weisen die Untersuchungsergebnisse darauf hin, dass die Prüfungspraxis nur beim Erhalt einer korrekten Mandantenerklärung effektiv ist (Trefferquote bei 92%). Ist die Mandantenerklärung hingegen inkorrekt oder holt sich der Prüfer keine Mandantenerklärung ein, so deuteten die Ergebnisse darauf hin, dass Prüfer Schwierigkeiten haben, die analytische Prüfungshandlung erfolgreich abzuschließen (in beiden Fällen lag die Trefferquote bei lediglich 36%).

#### **3.5.3.3.2.1.4.2 Einfluss von branchenspezifischer Erfahrung**

Eine Studie, die sich mit dem Einfluss von branchenspezifischer Erfahrung auf den Erfolg bei der Durchführung von analytischen Prüfungshandlungen beschäftigt, führte GREEN durch. Sie ließ insgesamt 58 Prüfer eine analytische Prüfungshandlung bearbeiten, bei der eine fehlerhafte Gemeinkostenallokation zu einer wesentlichen unerwarteten Abweichung in der Bruttogewinnmarge führte.<sup>2</sup> Bei 50% der Teilnehmer (n=29) handelte es sich um Branchenspezialisten. Die anderen 50% (n=29) wiesen zwar den gleichen Umfang an allgemeiner Prüfungserfahrung auf, verfügten jedoch über wesentlich weniger branchenspezifische Erfahrung als die Spezialisten.<sup>3</sup> Allen Teilnehmern wurde eine inkorrekte Mandantenerklärung mitgeteilt.<sup>4</sup>

---

<sup>1</sup> Für eine derartige Annahme spricht vor allem, dass Unternehmen der gleichen Branche in der Regel vergleichbare Strukturen und Relationen im Jahresabschluss aufweisen. Daher kann durchaus davon ausgegangen werden, dass Prüfer einer bestimmten Branche häufig ähnliche Bereiche bzw. Domänen prüfen.

<sup>2</sup> Vgl. GREEN (2004b), S. 9.

<sup>3</sup> Die Klassifizierung nach Branchen- bzw. Nicht-Branchenspezialist erfolgte nach einer Selbsteinschätzung der Prüfer. Diese Selbsteinschätzung konnte jedoch auch über deskriptiv-statistische Merkmale bestätigt werden. Die Spezialisten verfügten u.a über wesentlich mehr branchenspezifische Mandanten sowie Prüfungsaufträge. Vgl. GREEN (2004b), S. 9-10.

<sup>4</sup> Vgl. GREEN (2004b), S.11.



Im Ergebnis konnte GREEN feststellen, dass zwischen den Spezialisten und den Nicht-Spezialisten keine signifikanten Unterschiede hinsichtlich der Anzahl der selbstgenerierten plausiblen Fehler- bzw. Nicht-Fehlerhypothesen vorlagen. Spezialisten bzw. Nicht-Spezialisten generierten durchschnittlich 2,31 bzw. 2,66 Fehlerhypothesen und 3,59 bzw. 3,07 Nicht-Fehlerhypothesen. Darüber hinaus waren Spezialisten nicht wesentlich häufiger dazu in der Lage, die richtige Fehlerhypothese zu generieren (58,6% der Spezialisten (n=17) ggü. 51,7% der Nicht-Spezialisten (n=15)).<sup>1, 2</sup> Allerdings ließen die Ergebnisse auch darauf schließen, dass es den 17 Spezialisten, die die richtige Fehlerhypothese generierten, im Vergleich zu den 15 Nicht-Spezialisten, die ebenfalls die richtige Fehlerhypothese generierten, wesentlich häufiger möglich war, die analytische Prüfungshandlung auch erfolgreich abzuschließen. Von den 15 Nicht-Spezialisten wählten lediglich zwei (13,3%) die richtige Fehlerursache aus. Bei den 17 Spezialisten waren es hingegen neun (52,9%).<sup>3</sup>

Ähnlich wie bei der Studie von BEDARD/BIGGS legen auch die Ergebnisse der Studie von GREEN nahe, dass die in der Prüfungspraxis verstärkt zu beobachtende Branchenspezialisierung durchaus einen sinnvollen Beitrag zur Verbesserung der Prüfungsqualität leisten kann.<sup>4</sup>

#### **3.5.3.3.2.2 Der Einfluss des Erhalts einer Hypothese durch den Vorgesetzten auf die nachfolgende interne Hypothesengenerierung**

In der Prüfungspraxis kommt neben dem Mandanten insbesondere auch der Vorgesetzte (i.d.R. der Manager) als potenzielle Quelle von „vererbten“ Hypothesen in Betracht.<sup>5</sup> Es ist davon auszugehen, dass Prüfer die Meinung von einem direkten Vorgesetzten in einem anderen Lichte betrachten als Erklärungen vom Mandanten. Hiefür spricht etwa die bei einem Manager anzunehmende höhere Objektivität im Vergleich zu Mitarbeitern des zu prüfenden Unternehmens.<sup>6</sup>

---

<sup>1</sup> Vgl. GREEN (2004b), S. 12, 25.

<sup>2</sup> Da das zugrunde liegende Untersuchungsdesign jedoch keine Kontrollgruppe berücksichtigte, kann an dieser Stelle nicht beurteilt werden, ob vergleichbare Prüfer wesentlich mehr Fehlerhypothesen generiert hätten, wenn ihnen keine inkorrekte Mandantenerklärung bereitgestellt wurde. Insofern kann die (Nicht-)Existenz eines interference effects nicht beurteilt werden.

<sup>3</sup> Vgl. GREEN (2004b), S. 15.

<sup>4</sup> So auch GREEN (2004b), S. 17.

<sup>5</sup> Vgl. CHURCH/SCHNEIDER (1993), S. 334.

<sup>6</sup> Vgl. CHURCH/SCHNEIDER (1993), S. 334 m.w.N.

Da bei Mitarbeitern des Mandanten Anreize bestehen könnten, Fehler zu verbergen und dem Prüfer plausible, jedoch nicht den Tatsachen entsprechende Gründe für die Abweichungen zu liefern<sup>1</sup>, konzentrierte sich die Prüfungsforschung beim Erhalt einer Mandantenerklärung auf den Einfluss von Nicht-Fehlerhypothesen auf die eigene Hypothesengenerierung. Bei Vorgesetzten hingegen ist jedoch auch denkbar, dass sie auf potenzielle Fehlerursachen hinweisen. Daher hat sich die empirische Prüfungsforschung in diesem Bereich auf den Einfluss von Fehlerhypothesen auf die nachfolgende interne Hypothesengenerierung konzentriert.

In diesem Zusammenhang ist die Studie von LIBBY zu sehen. Er vermutete, dass die Weitergabe einer Fehlererklärung durch den Vorgesetzten analog zur Selbstgenerierung zu einer vermehrten Generierung von weiteren Fehlerhypothesen führt (spreading effect).<sup>2</sup> Die Ergebnisse ließen einen derartigen Effekt jedoch nicht erkennen.<sup>3</sup> LIBBY sowie LIBBY/FREDERICK vermuteten, dass das Ausbleiben des Effekts auf die Typizität des „vererbten“ Fehlers zurückzuführen sei. Eine theoretische Grundlage für diese Argumentation bietet *Rosch's theory of categorization*<sup>4</sup>, wonach „typical (frequent) category members do not activate other category members as strongly as nontypical ones.“<sup>5</sup> In LIBBYs Studie wurde den teilnehmenden Prüfern mitgeteilt, dass der Vorgesetzte als Fehlerursache die Einbuchung einer Forderung aus der nächsten Periode in der abgelaufenen Periode (sog. cut off-Fehler) vorschlägt.<sup>6</sup> Empirisch-archivistische Untersuchungen haben ergeben, dass es sich hierbei um einen häufig auftretenden und somit typischen Fehler handelt.<sup>7</sup> Daher sei verständlich, weshalb der vermutete spreading effect nicht eingetreten ist.<sup>8</sup>

Um diesen Zusammenhang genauer zu untersuchen, führten LIBBY/FREDERICK eine weitere Studie durch, bei der zwischen der „Vererbung“ einer typischen und atypischen Fehlerhypothese differenziert wurde.<sup>9</sup> Wie von den Autoren angenommen wurde, generierten Prüfer, die einen atypischen Fehler erhielten, signifikant mehr Fehlerhypothesen aus dem gleichen Transaktionskreis als Prüfer, denen keine Hypothese „vererbt“ wurde.

---

<sup>1</sup> An dieser Stelle sei auch auf die Möglichkeit der Verdeckung von Bilanzmanipulationen hingewiesen.

<sup>2</sup> Vgl. LIBBY (1985), S. 653 sowie Kap. 3.5.3.3.1.2.1.

<sup>3</sup> Vgl. LIBBY (1985), S. 661.

<sup>4</sup> Vgl. MERVIS/ROSCH (1981), S. 97.

<sup>5</sup> LIBBY (1985), S. 664.

<sup>6</sup> Vgl. LIBBY (1985), S. 665.

<sup>7</sup> Vgl. HYLAS/ASHTON (1982), S. 751-765; COACKLEY/LOEBBECKE (1985), S. 199-245.

<sup>8</sup> Vgl. LIBBY/FREDERICK (1990), S. 363.

<sup>9</sup> Vgl. LIBBY/FREDERICK (1990), S. 355. Eine ausführliche Beschreibung der Aufgabenstellung ist dem Kap. 3.5.3.3.1.2.1 zu entnehmen.

Hingegen konnte kein signifikanter Unterschied zwischen Prüfern, die einen typischen Fehler erhielten und Prüfern, denen kein Fehler zur Verfügung gestellt wurde, festgestellt werden.<sup>1</sup> Nach Ansicht der Autoren konnte demnach die Gültigkeit von *Rosch's theory of categorization* bestätigt und das Ausbleiben des spreading effects in LIBBYS Studie erklärt werden.

Kritisch zu beurteilen ist allerdings, dass sich die Studie von LIBBY/FREDERICK in einem entscheidenden Punkt von LIBBYS Studie unterscheidet und die Übertragbarkeit ihrer Ergebnisse somit in Frage gestellt werden kann. Während die Probanden bei LIBBY deutlich darauf hingewiesen wurden, dass es sich um eine Erklärung eines Vorgesetzten handelte, ging bei LIBBY/FREDERICK die Quelle der „vererbten“ Hypothese nicht aus der Aufgabenstellung hervor.<sup>2</sup> Somit ist nicht sichergestellt, dass die teilnehmenden Prüfer die Fehlerhypothese als eine Erklärung des Vorgesetzten wahrgenommen haben.

Eine Studie, die ebenfalls zwischen typischen und atypischen Fehlerhypothesen differenzierte führten CHURCH/SCHNEIDER durch.<sup>3</sup> Sie ließen insgesamt 96 Prüfer mit einer durchschnittlichen Prüfungserfahrung von etwa zwei Jahren einen Fall bearbeiten, bei dem eine wesentliche Abweichung in der Bruttogewinnmarge festzustellen war. Den Teilnehmern wurde mitgeteilt, dass sich der diese Abweichung verursachende Fehler entweder im Transaktionskreis Forderungen & Umsätze oder Verbindlichkeiten & Einkäufe befindet. Anschließend wurden die teilnehmenden Prüfer danach unterteilt, ob sie eine typische (atypische) Fehlerhypothese aus dem Transaktionskreis Forderungen & Umsätze bzw. eine typische (atypische) Fehlerhypothese aus dem Transaktionskreis Verbindlichkeiten & Einkäufe zur Verfügung gestellt bekamen. Damit die teilnehmenden Prüfer eindeutig erkennen konnten, dass es sich dabei um eine Meinung des Vorgesetzten handelte, wurden die Namen der Vorgesetzten, mit denen die jeweiligen Prüfer in der Prüfungspraxis tatsächlich zusammengearbeitet haben, aufgenommen. Einer weiteren Gruppe wurde keine Fehlerhypothese „vererbt“.

Im Ergebnis stellten CHURCH/SCHNEIDER fest, dass weder die „Vererbung“ von typischen noch von atypischen Fehlerhypothesen zu einer verstärkten Generierung von Fehlerhypothesen aus dem gleichen Transaktionskreis führten.<sup>4</sup> Insofern konnten die Ergebnisse von LIBBY/FREDERICK in dieser Studie nicht bestätigt werden. Vielmehr deuten

---

<sup>1</sup> Vgl. LIBBY/FREDERICK (1990), S. 360-361.

<sup>2</sup> Vgl. LIBBY/FREDERICK (1990), S. 364.

<sup>3</sup> Vgl. im Folgenden CHURCH/SCHNEIDER (1993), S. 338-339, 344-345.

<sup>4</sup> Vgl. CHURCH/SCHNEIDER (1993), S. 341-343, 345.

die Ergebnisse darauf hin, dass die teilnehmenden Prüfer Schwierigkeiten hatten, Fehlerhypothesen zu generieren, die aus dem gleichen transaction cycle stammten wie die „vererbte“ Hypothese.<sup>1</sup> Daher gehen CHURCH/SCHNEIDER davon aus, dass die Meinung eines Vorgesetzten eher zu einem interference effect führe als, wie von LIBBY bzw. LIBBY/FREDERICK vermutet, zu einem spreading effect. Den Beobachtungen zu Folge geht der interference effect sogar so weit, dass die teilnehmenden Prüfer eher dazu in der Lage waren, Fehlerhypothesen abzurufen, die nicht dem Transaktionskreis der „vererbten“ Hypothese zuzurechnen sind. CHURCH/SCHNEIDER bezeichnen diesen Effekt daher auch als „interference effect within a category.“<sup>2</sup>

### **3.5.3.3.2.3 Der Einfluss des Erhalts einer Hypothese durch Entscheidungshilfen**

#### **3.5.3.3.2.3.1 Vorbemerkungen**

Neben Mandanten bzw. dessen Mitarbeitern sowie Vorgesetzten kommen insbesondere auch Entscheidungshilfen als potenzielle Quelle einer externen Hypothese in Betracht. Bezogen auf analytische Prüfungshandlungen stellen Entscheidungshilfen dem Prüfer mögliche Hypothesen zur Verfügung, die für eine festgestellte unerwartete Abweichung in Betracht gezogen werden könnten. In der Regel können Entscheidungshilfen ohne erheblichen technischen Aufwand in die Prüfungssoftware der Prüfungsgesellschaften implementiert werden.

Der Einsatz von technischen Entscheidungshilfen ist vor allem vor dem Hintergrund der in unterschiedlichen Bereichen empirisch nachgewiesenen Phänomene des interference effects bzw. der truncated hypothesis generation gerechtfertigt.<sup>3</sup> Sollten Prüfer tatsächlich auf Schwierigkeiten stoßen, weitere potenzielle Fehlerursachen abzurufen, nachdem sie bereits Nicht-Fehlerhypothesen intern oder extern generiert haben, so stellen Entscheidungshilfen eine sinnvolle Alternative dar, um diesen Verzerrungen entgegenzuwirken.

---

<sup>1</sup> Vgl. CHURCH/SCHNEIDER (1993), S. 341.

<sup>2</sup> CHURCH/SCHNEIDER (1993), S. 337.

<sup>3</sup> Für den interference effect im Rahmen der internen Hypothesengenerierung sei auf Kap. 3.5.3.3.1.2.2 verwiesen. In Bezug auf den interference effect bzw. der hiermit unmittelbar in Zusammenhang stehenden truncated hypothesis generation soll auf Kap. 3.5.3.3.2.1.3.2 verwiesen werden.

Die Verwendung von Entscheidungshilfen ermöglicht es dem Prüfer, eine Vielzahl von möglichen Gründen für eine unerwartete Abweichung zu berücksichtigen.<sup>1</sup> Aus Effektivitätsgründen wäre eine hohe Anzahl der vom Prüfer in Betracht gezogenen plausiblen Hypothesen positiv zu beurteilen, da sich die Wahrscheinlichkeit der Berücksichtigung des tatsächlichen Fehlers mit der zunehmenden Anzahl plausibler Abweichungsgründe erhöht. Andererseits bedarf die Überprüfung einer Vielzahl von Hypothesen eines enormen zeitlichen Aufwands, so dass hieraus Beeinträchtigungen der Effizienz zu erwarten sind. Darüber hinaus kann die gleichzeitige Überprüfung von zu vielen Hypothesen zu einer Überbelastung der mentalen Kapazität des jeweiligen Prüfers führen (sog. *cognitive overload*).<sup>2</sup> Aus diesen Gründen wird der Prüfer um die Reduzierung der vorgeschlagenen Hypothesen bemüht sein.

#### **3.5.3.3.2.2 Der Einfluss der Auswahlstrategie auf die Anzahl der ausgewählten Hypothesen**

Bei der Reduzierung einer Vielzahl von Alternativen auf ein überschaubares Maß kommen zwei unterschiedliche Strategien in Betracht. Bei der ersten Strategie werden alle Alternativen, von denen das Individuum glaubt, dass sie nicht als der tatsächliche Grund in Betracht kommen, nicht weiter berücksichtigt (sog. Exklusion).<sup>3</sup> Bei der zweiten Strategie geht das Individuum umgekehrt vor. Hier werden ausschließlich diejenigen Alternativen weiterhin berücksichtigt, die das Individuum als tatsächlichen Grund in Betracht zieht (sog. Inklusion).<sup>4</sup> Unabhängig davon, welche Strategie angewandt wird, sollten beide zu dem gleichen Ergebnis kommen. Allerdings konnte in Studien aus der psychologischen Forschung nachgewiesen werden, dass die Verwendung der Inklusion (Exklusion) zu einer geringeren (höheren) Anzahl weiter betrachteter Alternativen führt.<sup>5</sup> Der Grund hierfür wird in der Verwendung unterschiedlicher Wahrscheinlichkeitsbandbreiten gesehen. Bei der Inklusion wird die untere Wahrscheinlichkeitsgrenze für den Verbleib einer Hypothese niedriger gesetzt (z.B. 50-100%) als die obere Wahr-

---

<sup>1</sup> In der empirischen Prüfungsforschung werden den Probanden bei Studien zur Untersuchung des Einflusses von Entscheidungshilfen in der Regel 15-20 potenzielle Gründe zur Verfügung gestellt. Vgl. etwa ANDERSON/JENNINGS/KAPLAN/RECKERS (1995), S. 33-62; ANDERSON/KAPLAN/RECKERS (1997), S. 1-20.

<sup>2</sup> Vgl. etwa BHATTACHARJEE ET AL. (1999), S. 98 sowie Kap. 3.5.3.3.1.3.2.

<sup>3</sup> In der Regel werden die unwahrscheinlichen Alternativen auf der Liste durchgestrichen.

<sup>4</sup> Etwa durch Markierung oder Hervorhebung der wahrscheinlichen Alternativen.

<sup>5</sup> Vgl. LEVIN/JASPER/FORBES (1998), S. 193-210; YANIV/SCHUL (2000), S. 293-313.

scheinlichkeitsgrenze für die Eliminierung einer Alternative bei Verwendung der Exklusion (etwa 0-60%).<sup>1</sup>

Um zu untersuchen, ob diese Auswirkungen des sog. *goal framings* auch im Kontext von analytischen Prüfungshandlungen zu beobachten sind, ließen MUELLER/ANDERSON insgesamt 65 Prüfer mit einer durchschnittlichen Prüfungserfahrung von etwa 5,5 Jahren eine Fallstudie bearbeiten, bei der eine wesentliche Reduzierung der Umschlagshäufigkeit des abgelaufenen Geschäftsjahrs im Vergleich zum Vorjahr festgestellt werden konnte. Die teilnehmenden Prüfer erhielten eine Liste mit 20 in Frage kommenden Abweichungsursachen und wurden aufgefordert, eine engere Auswahl potenzieller Fehlerursachen zu treffen. Dabei wurden die Prüfer aufgefordert, entweder diejenigen Ursachen zu markieren, die sie für wahrscheinlich (inclusion frame) oder für unwahrscheinlich (exclusion frame) hielten.<sup>2</sup> Im Hinblick auf die Möglichkeit eines debiasings wurde zusätzlich untersucht, ob die möglichen Auswirkungen des goal framings durch ein erhöhtes Prüfungsrisiko gemindert oder beseitigt werden können.<sup>3</sup>

Im Ergebnis konnten die Autoren feststellen, dass die Inklusion eine größere Reduzierung der Ausgangsliste zur Folge hatte als die Exklusion. Demnach führte die Inklusions-Strategie zu signifikant weniger auserwählten Abweichungshypothesen als die Exklusions-Strategie (6,7 ggü. 12,2).<sup>4</sup> Darüber hinaus konnte herausgefunden werden, dass ein erhöhtes Prüfungsrisiko den Auswirkungen des goal framings nicht entgegenwirken konnte. Zwar umfassten die reduzierten Listen insgesamt mehr potenzielle Abweichungsursachen, was für eine gesteigerte prüferische Skepsis spricht. Dennoch führte goal framing auch hier zu signifikanten Unterschieden in den reduzierten Hypothesensätzen (9,2 ggü. 14,8).<sup>1</sup>

Insgesamt kann also festgehalten werden, dass mit goal framing eine weitere potenzielle Verzerrung bei der Durchführung von analytischen Prüfungshandlungen empirisch nachgewiesen werden konnte. Ein inclusion frame (exclusion frame) führte dazu, dass sowohl bei niedrigem als auch bei hohem Prüfungsrisiko 5,5 Hypothesen weniger (mehr) ausgewählt wurden. Inwiefern sich dieses auf die Effektivität und Effizienz von Abschlussprüfungen auswirkt, kann lediglich bedingt beantwortet werden. Eine höhere Anzahl von potenziellen Gründen könnte zu einer höheren Effektivität führen, da eine

---

<sup>1</sup> Vgl. YANIV/SCHUL (1997), S. 216-218.

<sup>2</sup> Vgl. MUELLER/ANDERSON (2002), S. 165-168.

<sup>3</sup> Vgl. MUELLER/ANDERSON (2002), S. 163-165.

<sup>4</sup> Vgl. MUELLER/ANDERSON (2002), S. 170.

größere Anzahl potenzieller Hypothesen überprüft werden. Andererseits kann eine geringere Anzahl von Gründen bei gleicher Effektivität zu einer höheren Effizienz führen.<sup>2</sup>

Eine Handlungsempfehlung für die Prüfungspraxis sollte sich daher in erster Linie an einer möglichst „unverzerrten“ prüferischen Urteilsfähigkeit orientieren. Bei der Implementierung von Entscheidungshilfen sollten Wirtschaftsprüfungsunternehmen darauf achten, dass goal framing nach Möglichkeit vermieden wird. Darüber hinaus sollten Prüfern Strategien vermittelt werden, die es ihnen im Umgang mit Entscheidungshilfen ermöglichen, eine möglichst unbeeinflusste Hypothesenauswahl zu treffen. Zu denken wäre hier etwa an eine Quantifizierung der Eintrittswahrscheinlichkeiten i.V.m. einer fest vorgegebenen Wahrscheinlichkeitsbandbreite für die Inklusion oder Exklusion von Hypothesen. Diese Bandbreiten könnten dann in Abhängigkeit von dem geschätzten Prüfungsrisiko variiert werden.

### **3.5.3.4 Der Einfluss von Gruppenprozessen auf die Hypothesengenerierung**

#### **3.5.3.4.1 Vorbemerkungen**

Ein Sonderfall der Hypothesengenerierung liegt vor, wenn die Abweichungshypothesen nicht durch eine Person abgerufen wurden, sondern eine Gruppe von Personen gemeinsam potenzielle Gründe für die festgestellte Abweichung erarbeitet. Diese Gruppenprozesse stellen eine Mischform der internen und externen Generierung dar, da sowohl ein interner Hypothesenabruf durch einzelne Gruppenmitglieder als auch eine Mitteilung weiterer Hypothesen durch andere Gruppenmitglieder erfolgt.

Die Betrachtung von Gruppenprozessen berücksichtigt insbesondere, dass die Abschlussprüfung in der Regel in Prüfungsteams durchgeführt wird und somit einen multipersonellen Charakter hat.<sup>3</sup> Innerhalb der Teammitglieder kann es häufig zu Gruppen- bzw. Interaktionsprozessen kommen.<sup>1</sup>

Bei diesen Gruppen- bzw. Kommunikationsprozessen kann differenziert werden zwischen der Kommunikation der Teammitglieder untereinander, etwa beim Auftauchen von Problemen, zu dessen Lösung das einzelne Teammitglied die Meinungen anderer

---

<sup>1</sup> Vgl. MUELLER/ANDERSON (2002), S. 170.

<sup>2</sup> Vgl. Kap. 3.5.3.3.1.3.2.

<sup>3</sup> Vgl. SCHREIBER (2000), S. 103-106 oder RICHTER (1999), S. 267.

Teammitglieder einholen möchte, und der Kommunikation zwischen einzelnen Teammitgliedern und dem Manager im Rahmen des Review-Prozesses. Bei Erstgenannter handelt es sich in der Regel um Diskussionen, bei denen Informationen gegenseitig ausgetauscht werden. Bei Letztgenannter ist zu berücksichtigen, dass es nicht immer zu einer direkten Diskussion („face-to-face“) zwischen dem Teammitglied und dem Reviewer kommen muss. Denkbar wäre auch, dass der Manager die Arbeitspapiere und Prüfungsfeststellungen in Abwesenheit des erstellenden Prüfers durchschaut.<sup>2</sup> Dennoch handelt es sich auch hierbei um eine Form der Kommunikation, da hier Informationen von einem Individuum (Teammitglied) an ein anderes Individuum (Manager) schriftlich weitergegeben werden. Im Gegensatz zu einer direkten Diskussion fließen Informationen hier allerdings nur in eine Richtung.

#### **3.5.3.4.2 Der Einfluss des Review-Prozesses auf die Anzahl der plausiblen Hypothesen**

ISMAIL/TROTMAN untersuchten den Einfluss von Gruppenprozessen auf die Quantität und Qualität der generierten Hypothesen. Sie vermuteten, dass Prozesse zwischen dem Manager und den Prüfern eine sog. *cognitive interstimulation* bewirken, die dazu führen, dass mehr plausible Hypothesen generiert werden.<sup>3</sup> Darüber hinaus könnten Gruppenprozesse auch dazu führen, dass mehr falsche Schlussfolgerungen und somit inkorrekte Hypothesen korrigiert werden (sog. *error correction*).<sup>4</sup> Sowohl cognitive interstimulation als auch error correction konnten in der psychologischen Forschung als Grund für die verbesserte Hypothesengenerierung von Gruppen nachgewiesen werden.<sup>5</sup>

Um festzustellen, ob derartige Phänomene auch im Kontext von analytischen Prüfungshandlungen von Relevanz sind, bildeten ISMAIL/TROTMAN 38 Gruppen, bestehend aus jeweils einem Prüfer und zwei Reviewern.<sup>6</sup> Die Aufgabe des Prüfers bestand darin, potenzielle Fehlerhypothesen für eine festgestellte wesentliche Abweichung zu generie-

---

<sup>1</sup> Vgl. SCHREIBER (2000), S. 106 oder RICHTER (1999), S. 274-280.

<sup>2</sup> Vgl. ISMAIL/TROTMAN (1995), S. 346.

<sup>3</sup> Vgl. ISMAIL/TROTMAN (1995), S. 347. Als Grund hierfür führen ISMAIL/TROTMAN den facilitation effect auf. Letztendlich handelt es sich hierbei um ein Synonym für den spreading effect, welcher unter Kap. 3.5.3.3.1.2.1 ausführlich erläutert wurde.

<sup>4</sup> Vgl. ISMAIL/TROTMAN (1995), S. 347.

<sup>5</sup> Vgl. HACKMAN/MORRIS (1975), S. 45-99; HILL (1982), S. 517-539; HASTIE/PENROD/PENNINGTON (1983); HINSZ (1990), 705- 718.

<sup>6</sup> Die 38 Prüfer wiesen eine durchschnittliche Prüfungserfahrung von etwa 1,8 Jahren auf. Bei den Reviewern handelte es sich um 42 Manager mit einer durchschnittlichen Prüfungserfahrung von 6,3 Jahren und 34 Prüfungsleiter mit einer durchschnittlichen Prüfungserfahrung von 3,4 Jahren.



ren.<sup>1</sup> Im anschließenden Review-Prozess durch einen der Reviewer waren die Prüfer weiterhin anwesend, so dass die Möglichkeit zur direkten Diskussion bestand. Anschließend wurden die ursprünglich vom Prüfer generierten Hypothesensätze erneut durch den zweiten Reviewer gesichtet, allerdings in Abwesenheit des Prüfers, so dass keine Diskussionen stattfinden konnten.<sup>2</sup>

Im Ergebnis konnten die Autoren feststellen, dass der Review-Prozess zu einer signifikant höheren Anzahl von plausiblen Fehlerhypothesen führt. Die Prüfer generierten im Schnitt etwa 5,45 plausible Hypothesen. Hingegen wurden nach dem Review-Prozess durchschnittlich 7,45 generiert.<sup>3</sup> ISMAIL/TROTMAN vermuteten, dass die erhöhte Zahl dem cognitive interstimulation zugerechnet werden kann.

Um zusätzlich festzustellen, welche Art von Kommunikation welchen Einfluss auf die Anzahl der plausiblen Hypothesen hatte, differenzierten die Autoren zwischen Gruppenprozessen ohne und mit Diskussion. Dabei zeigte sich, dass der Review-Prozess ohne Diskussion zu 7,0 generierten Hypothesen führte und der Review-Prozess mit Diskussion zu 7,9.<sup>4</sup> Es zeigt sich demnach, dass der durch den Review- bzw. Gruppenprozess gewonnene Nutzen eine weitere Erhöhung durch die Möglichkeit von direkten Diskussionen erfuhr. Interessanterweise konnten die Autoren durch weitere Analysen identifizieren, dass die höhere Anzahl der plausiblen Hypothesen bei den Gruppen mit Diskussion nicht darin begründet war, dass weitere plausible Hypothesen hinzugefügt wurden. Vielmehr zeichneten sich diese Gruppen dadurch aus, dass sie weniger plausible Hypothesen verwarfen als Gruppen ohne Diskussion.<sup>5</sup> Dieses könnte in erster Linie darin begründet sein, dass die Prüfer ihre Hypothesen im direkten Gespräch mit dem Reviewer rechtfertigen und ihn von der Angemessenheit ihrer Hypothesen überzeugen konnten, so dass der Reviewer von einer Eliminierung der jeweiligen Hypothese abgesehen hatte.

Nicht bestätigen konnten die Autoren hingegen ihre Vermutung, dass Gruppen den Individuen überlegen sind, weil sie gemeinsam mehr Fehler korrigieren. Dieses begrün-

---

<sup>1</sup> Vgl. ISMAIL/TROTMAN (1995), S. 348. Da die Autoren die Aufgabenstellung von LIBBY/FREDERICK (1990) übernommen haben, sei an dieser Stelle auf die ausführlich Darstellung der LIBBY/FREDERICK-Studie unter Kap. 3.5.3.3.1.1.1 verwiesen.

<sup>2</sup> Vgl. ISMAIL/TROTMAN (1995), 348-349.

<sup>3</sup> Vgl. ISMAIL/TROTMAN (1995), S. 351 i.V.m. 352.

<sup>4</sup> Vgl. ISMAIL/TROTMAN (1995), S. 351.

<sup>5</sup> Vgl. ISMAIL/TROTMAN (1995), S. 353.

den die Autoren jedoch damit, dass die teilnehmenden Gruppen insgesamt wenig inkorrekte Hypothesen generierten, die anschließend hätten korrigiert werden können.<sup>1</sup>

#### **3.5.3.4.3 Der Einfluss der Kommunikation unter den Mitgliedern eines Prüfungsteams auf die Anzahl der plausiblen Hypothesen und den Erfolg der analytischen Prüfungshandlung**

BEDARD/BIGGS/MARONEY untersuchten in ihrer Studie die Kommunikation innerhalb des Prüfungsteams, also unter gleichberechtigten Teammitgliedern, wobei ausschließlich die direkte Kommunikation in Betracht gezogen wurde. Die Autoren beriefen sich in erster Linie auf das Modell von STEINER, wonach Gruppenprozesse nicht ausschließlich mit Nutzegewinnen verbunden sind, sondern auch Nutzenverluste verursachen können.<sup>2</sup> STEINER geht davon aus, dass Nutzegewinne einerseits durch *knowledge pooling* (d.h., dass die einzelnen Gruppenmitglieder ihr Wissen bündeln) und andererseits durch *error correction* entstehen können.<sup>3</sup> Als Ursachen für Nutzenverluste kommt einerseits die Dominanz einzelner Gruppenmitglieder in Betracht, wodurch andere Gruppenmitglieder ihr relevantes Wissen möglicherweise nicht einbringen können. Andererseits könnten sich die in Gruppenprozessen tendenziell häufig vorkommenden Störungen negativ auswirken, da sie dazu führen könnten, dass wichtige Gedankengänge nicht beendet werden.<sup>4</sup>

BEDARD/BIGGS/MARONEY ließen insgesamt acht Gruppen bestehend aus jeweils drei Prüfern<sup>5</sup> eine analytische Prüfungshandlung durchführen.<sup>6</sup> Um auf Unterschiede zu der Vorgehensweise einzelner Prüfer eingehen zu können, verglichen die Autoren die Gruppenergebnisse mit den Ergebnissen aus der Studie von BEDARD/BIGGS, die die gleiche Fallstudie von elf einzelnen Prüfern bearbeiten ließen.<sup>7</sup> Im Ergebnis konnten die

---

<sup>1</sup> Vgl. ISMAIL/TROTMAN (1995), S. 352-353.

<sup>2</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 207.

<sup>3</sup> Eine Beschreibung von *error correction* erfolgte bereits unter Kap. 3.5.3.4.2.

<sup>4</sup> Vgl. STEINER (1972).

<sup>5</sup> BEDARD/BIGGS/MARONEY machen keine expliziten Angaben zu der durchschnittlichen Prüfungserfahrung der Teilnehmer. Es wurde lediglich darauf hingewiesen, dass es sich um Prüfungsleiter handelte. Generell kann davon ausgegangen werden, dass Prüfungsleiter durchschnittlich zwei bis drei Jahre Berufserfahrung aufweisen.

<sup>6</sup> Da es sich hierbei um die gleiche Aufgabenstellung wie bei BEDARD/BIGGS (1991a) handelt (BEDARD/BIGGS/MARONEY (1998), S. 217), sei für eine ausführliche Fallbeschreibung auf Kap. 3.5.2.1 verwiesen.

<sup>7</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 217. Demnach kann davon ausgegangen werden, dass die teilnehmenden Prüfer etwa über den gleichen Umfang an Prüfungserfahrung verfügten wie die Prüfer aus der BEDARD/BIGGS-Studie.

Autoren erkennen, dass die Prüfergruppen signifikant häufiger die richtige Fehlerhypothese generierten als die einzelnen Prüfer. Insgesamt waren es fünf von acht Gruppen (62,5%) im Vergleich zu einem von elf Einzelprüfern (9,1%).<sup>1</sup> Die Autoren führten diese Unterschiede im Wesentlichen darauf zurück, dass die Teammitglieder dazu fähig waren, ihr Wissen zu bündeln (*knowledge pooling*). Sie zogen aus den Diskrepanzen in den Jahresabschlussdaten wesentlich häufiger die richtigen Schlussfolgerungen als die Einzelprüfer und generierten mehr plausible Hypothesen.<sup>2</sup> In Bezug auf error correction konnten hingegen keine signifikanten Unterschiede zwischen Gruppen und Einzelprüfern festgestellt werden. Beide korrigierten begangene Fehler in 67% aller Fälle.<sup>3</sup>

Der Vorteil des knowledge pooling in der Phase der Hypothesengenerierung konnte in der weiteren Vorgehensweise jedoch nicht aufrechterhalten werden. Lediglich zwei von den fünf Gruppen mit erfolgreicher Hypothesengenerierung waren dazu in der Lage, die analytische Prüfungshandlung erfolgreich abzuschließen. Demgegenüber wählte der einzige Einzelprüfer, dessen Hypothesensatz die richtige Hypothese umfasste, diese abschließend auch aus.<sup>4</sup> Die Autoren stellten im Wesentlichen zwei Ursachen für die mangelhafte Vorgehensweise der Gruppen fest. Zum einen wiesen die Gruppen, die die analytische Prüfungshandlung nicht erfolgreich abschließen konnten, auch die höchste Anzahl von Unterbrechungen (interruptions) auf, welche entweder dazu führten, dass eine richtige Vorgehensweise gestört und nicht weiter fortgesetzt wurde, oder dass eine fehlerhafte Vorgehensweise einsetzte. Die beiden einzigen Gruppen, welche die richtige Hypothese aus den diversen plausiblen Hypothesen selektierten, wiesen nicht eine einzige Unterbrechung auf.<sup>5</sup> Diese beiden Gruppen waren zudem als Einzige dazu in der Lage, ihre inkorrekten Fehlerhypothesen zu erkennen und zu korrigieren. Die Gruppen, die es nicht schafften, die korrekte Fehlerhypothese aus ihrem Hypothesensatz zu selektieren, waren hingegen am wenigsten dazu in der Lage, ihre inkorrekten Hypothesen zu korrigieren.<sup>6</sup>

Einen Nachweis für den negativen Einfluss von Dominanz einer Person konnten die Autoren hingegen nicht erbringen.<sup>7</sup> Im Gegenteil konnte sogar festgestellt werden, dass

---

<sup>1</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 217-218.

<sup>2</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 219-220.

<sup>3</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 220.

<sup>4</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 218.

<sup>5</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 223-225.

<sup>6</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 220-221.

<sup>7</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 223.

Dominanz insgesamt zu einer erfolgreichen Hypothesengenerierung und -selektion führte, wenn die dominierende Person durch eine hohe Kompetenz gekennzeichnet ist.<sup>1</sup>

Insgesamt betrachtet kann demnach festgehalten werden, dass die Gruppen den Einzelprüfern aufgrund der höheren Anzahl plausibler Hypothesen zwar überlegen waren. Allerdings waren sie aufgrund der gruppenbedingten Ineffizienzen nicht dazu in der Lage, dieses Erfolgspotenzial in der weiteren Vorgehensweise vollständig umzusetzen.<sup>2</sup> Gruppenmitgliedern sollten daher die Möglichkeit gegeben werden, ihre Gedankengänge vollständig zu erklären, da es ansonsten wenig nutzt, wenn sie eine plausible Hypothese generieren, die nicht von den anderen Gruppenmitgliedern akzeptiert wird.<sup>3</sup>

---

<sup>1</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 228-229.

<sup>2</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 226.

<sup>3</sup> Vgl. BEDARD/BIGGS/MARONEY (1998), S. 229.

### 3.5.3.5 Tabellarische Darstellung der empirischen Befunde zur Phase der Hypothesengenerierung

| Autor(en)                      | Teilnehmer |                                                     | Daten-<br>gewinnung | Daten-<br>auswertung | Ergebnisse                                                                                                                                                                                                                                                                                                           | Signifikanz |
|--------------------------------|------------|-----------------------------------------------------|---------------------|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
|                                | Anzahl     | Art                                                 |                     |                      |                                                                                                                                                                                                                                                                                                                      |             |
| BEDARD/BIGGS (1991a)           | 21         | 10 Manager<br>k.A.<br><br>11 Prüfungsleiter<br>k.A. | Experiment          | Verbale Protokolle   | Prüfer, deren anfängliche Hypothesengenerierung nicht die korrekte Abweichungsursache umfasst, haben im Rahmen der nachträglichen Hypothesengenerierung Probleme, die korrekte Abweichungsursache zu generieren.                                                                                                     | k.A.        |
| BIERSTAKER/BEDARD/BIGGS (1999) | 12         | 12 Prüfungsleiter<br>3 bis 5 J.                     | Experiment          | Verbale Protokolle   | Prüfer, deren anfängliche Hypothesengenerierung nicht die korrekte Abweichungsursache umfasst, haben im Rahmen der nachträglichen Hypothesengenerierung Probleme, die korrekte Abweichungsursache zu generieren.                                                                                                     | k.A.        |
| GREEN/TROTMAN (2003)           | 82         | k.A.<br>Ø = 3,8 J.                                  | Experiment          | Binomialtest         | Die Häufigkeit, mit der die richtige Hypothese nachträglich generiert wurde (n=13), ist signifikant größer als null. D.h., die anfängliche Generierung der richtigen Abweichungsursache ist keine notwendige Bedingung für den Erfolg von analytischen Prüfungshandlungen.                                           | 0,000       |
|                                |            |                                                     |                     | Binomialtest         | Die Häufigkeit, mit der Prüfer die richtige Hypothese zwar anfänglich generierten und diese jedoch nicht auswählten (n=39), ist signifikant größer als null. D.h., die anfängliche Generierung der richtigen Abweichungsursache ist keine hinreichende Bedingung für den Erfolg von analytischen Prüfungshandlungen. | 0,000       |
| GREEN (2004b)                  | 58         | k.A.<br>Ø = 4 J.                                    | Experiment          | Häufigkeiten         | Etwa 64% der Prüfer, die die analytische Prüfungshandlung erfolgreich abschlossen, generierten diese im Rahmen der nachträglichen Hypothesengenerierung.                                                                                                                                                             | k.A.        |
| ASARE/WRIGHT (2003)            | 102        | k.A.                                                | Experiment          | Häufigkeiten         | Von den 14 Prüfern, die die richtige Abwei-                                                                                                                                                                                                                                                                          | k.A.        |

|                                |     |                                                                    |            |                                      |                                                                                                                                                                                                                                                                                                                                                                |                       |
|--------------------------------|-----|--------------------------------------------------------------------|------------|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
|                                |     | Ø =2,05 J.                                                         |            |                                      | chungsursache nachträglich generierten, waren alle dazu in der Lage, die analytische Prüfungshandlung erfolgreich abzuschließen.                                                                                                                                                                                                                               |                       |
| LIBBY (1985)                   | 37  | Manager<br>Ø = 5 J.                                                | Experiment | Korrelationsanalyse<br>(Pearson)     | Es besteht ein starker Zusammenhang zwischen den generierten Hypothesen und den Häufigkeiten, mit denen sich Fehler im Jahresabschluss niederschlagen (sog. frequency knowledge).                                                                                                                                                                              | < 0,01                |
| LIBBY/FREDERICK (1990)         | 126 | 61 Manager<br>Median = 5 J.<br><br>70 Assistenten<br>Median = 1 J. | Experiment | Varianzanalyse                       | Manager generieren signifikant mehr Abweichungsursachen mit einer hohen Eintrittshäufigkeit als Prüfungsassistenten.                                                                                                                                                                                                                                           | 0,000                 |
| LIBBY (1985)                   | 37  | Manager<br>Ø = 5 J.                                                | Experiment | Chi-Quadrat-Test                     | Es besteht ein positiver Zusammenhang zwischen den erst kürzlich entdeckten Fehlern und der generierten Abweichungshypothesen (sog. recency effect).                                                                                                                                                                                                           | 0,004                 |
| LIBBY (1985)                   | 37  | Manager<br>Ø = 5 J.                                                | Experiment | t-Test                               | Es besteht ein signifikanter Zusammenhang zwischen dem Transaktionskreis der zuerst generierten Hypothese und dem Transaktionskreis der anschließend generierten Hypothesen.                                                                                                                                                                                   | 0,021                 |
| BEDARD/BIGGS (1991a)           | 21  | 10 Manager<br>k.A.<br><br>11 Prüfungsleiter<br>k.A.                | Experiment | Verbale Protokolle                   | Prüfer konzentrieren sich verstärkt auf den Transaktionskreis der zuerst generierten Hypothese und weniger auf Transaktionskreise der anschließend generierten Hypothesen.                                                                                                                                                                                     | k.A.                  |
| ANDERSON/KAPLAN/RECKERS (1992) | 38  | Manager<br>Ø = 5 J.                                                | Experiment | Varianzanalyse<br><br>Varianzanalyse | Prüfer, die zuerst Nicht-Fehlerhypothesen generierten, generierten anschließend signifikant weniger Fehler-Hypothesen, als Prüfer, die zuerst Fehler-Hypothesen generierten (sog. interference effect).<br><br>Prüfer, die zuerst Fehlerhypothesen generierten, generierten anschließend nicht signifikant weniger Nicht-Fehlerhypothesen, als Prüfer, die zu- | < 0,001<br><br>> 0,67 |

|                                 |     |                                                                    |            |                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                   |
|---------------------------------|-----|--------------------------------------------------------------------|------------|-------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|
|                                 |     |                                                                    |            |                                                             | erst Nicht-Fehlerhypothesen generierten (sog. asymmetric interference effect).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                   |
| LIBBY/FREDERICK (1990)          | 126 | 61 Manager<br>Median = 5 J.<br><br>70 Assistenten<br>Median = 1 J. | Experiment | Varianzanalyse<br><br>Varianzanalyse                        | Manager generieren signifikant mehr plausible Hypothesen als Prüfungsassistenten.<br>Manager generieren signifikant weniger nicht plausible Hypothesen als Assistenten.                                                                                                                                                                                                                                                                                                                                                                                                                                        | 0,000<br><br>0,000                                |
| ISMAIL/TROTMAN (1995)           | 38  | k.A.<br>$\bar{X} = 1,8$ J.                                         | Experiment | Mittelwertberechnung                                        | Prüfer generieren im Schnitt 5,4 plausible Hypothesen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | k.A.                                              |
| BHATTACHARJEE/KIDA/HANNO (1999) | 60  | k.A.<br>$\bar{X} = 3,7$                                            | Experiment | Mittelwertberechnung                                        | Prüfer generieren im Schnitt 5,0 plausible Hypothesen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | k.A.                                              |
| WRIGHT/WRIGHT (1997)            | 72  | Prüfungsleiter,<br>Manager, Partner<br>$\bar{X} = 8,41$            | Experiment | t-Test<br><br>t-Test                                        | Erfahrene Prüfer ( $\bar{X} \geq 12$ J.) generieren nicht signifikant mehr plausible Hypothesen als unerfahrene Prüfer ( $\bar{X} \leq 3$ J.)<br>Prüfer mit hoher Branchenerfahrung (mind. 7 Prüfungsaufträge in der Branche) generieren signifikant mehr plausible Hypothesen als Prüfer ohne Branchenerfahrung.                                                                                                                                                                                                                                                                                              | $\geq 0,10$<br><br>< 0,05                         |
| BHATTACHARJEE/KIDA/HANNO (1999) | 38  | k.A.<br>$\bar{X} = 1,8$ J.                                         | Experiment | Varianzanalyse<br><br>Kruskal-Wallace<br><br>Mann-Whitney-U | Prüfer, die anfänglich 3 Hypothesen generierten, benötigten für die Erledigung der Aufgabe signifikant weniger Zeit als Prüfer, die anfänglich eine Hypothese bzw. sechs Hypothesen generierten, und Prüfer, denen keine Anzahl der anfänglich zu generierenden Hypothesen vorgegeben wurde.<br>Die Anzahl der anfänglich generierten Hypothesen hat einen signifikanten Einfluss auf die Erfolgsquote der Prüfer.<br>Drei-Hypothesen-Prüfer sind nicht signifikant häufiger erfolgreich als Ein-Hypothesen-Prüfer, aber signifikant häufiger erfolgreich als Sechs-Hypothesen-Prüfer und Prüfer ohne Vorgabe. | < 0,048<br><br>0,036<br><br>0,164/0,007/<br>0,029 |
| BEDARD/BIGGS (1991a)            | 21  | 10 Manager                                                         |            | Verbale Protokolle                                          | Die mangelnde Fähigkeit eine anfänglich gene-                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | k.A.                                              |

|                                    |    |                                                                                           |            |                    |                                                                                                                                                                                                                                                                                                                                                |       |
|------------------------------------|----|-------------------------------------------------------------------------------------------|------------|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                                    |    | k.A.<br>11 Prüfungsleiter<br>k.A.                                                         | Experiment | Verbale Protokolle | rierte Hypothese zu wechseln, stellt den Hauptgrund für die schlechte Erfolgsquote dar. Prüfer sind selbst dann nicht in der Lage die Hypothese zu wechseln, wenn gegenteilige Beweise erlangt wurden.                                                                                                                                         | k.A.  |
| HEIMAN-HOFFMAN/MOSER/JOSEPH (1995) | 54 | k.A.<br>$\bar{d} = 3$ J.                                                                  | Experiment | Varianzanalyse     | Die Bereitschaft eine Hypothese zu wechseln hat einen signifikanten Einfluss auf die Erfolgsquote.                                                                                                                                                                                                                                             | 0,001 |
| HEIMAN (1990)                      | 73 | 26 Manager<br>$\bar{d} = 5,7$ J.<br><br>Prüfungsassistenten/-leiter<br>$\bar{d} = 2,8$ J. | Experiment | Varianzanalyse     | Die Vererbung einer Hypothese hat einen signifikanten Einfluss auf die Anzahl der anschließend selbst generierten Hypothesen (sog. interference effect).                                                                                                                                                                                       | 0,022 |
| NG/GREEN/SIMNETT (2001)            | 91 | k.A.<br>$\bar{d} = 3,1$ J.                                                                | Experiment | Varianzanalyse     | Prüfer mit Mandantenerklärung generieren anschließend nicht signifikant weniger Hypothesen als Prüfer ohne Mandantenerklärung.                                                                                                                                                                                                                 | 0,945 |
| GREEN (2004a)                      | 63 | k.A.<br>$\bar{d} = 3,84$ J.                                                               | Experiment | Varianzanalyse     | Prüfer, die eine Nicht-Fehlererklärung vom Mandanten erhalten, generieren anschließend nicht signifikant weniger Fehlerhypothesen, als Prüfer, die keine Mandantenerklärung erhalten.                                                                                                                                                          | 0,604 |
| GREEN (2004a)                      | 63 | k.A.<br>$\bar{d} = 3,84$ J.                                                               | Experiment | Varianzanalyse     | Prüfer, die eine Nicht-Fehlererklärung vom Mandanten erhalten, generieren anschließend signifikant mehr Nicht-Fehlerhypothesen, als Prüfer, die keine Mandantenerklärung erhalten (sog. spreading effect).                                                                                                                                     | 0,022 |
| BIERSTAKER/BEDARD/BIGGS (1999)     | 12 | 12 Prüfungsleiter<br>3 bis 5 J.                                                           | Experiment | Verbale Protokolle | Der Erhalt einer Mandantenerklärung führte bei 10 Prüfern dazu, dass sie trotz der Existenz von Hinweisen, die auf die Inkorrektheit der Mandantenerklärung hindeuteten, diese fortwährend evaluierten (sog. wheel spinning), und dass es ihnen nicht möglich war, eine alternative Abweichungsursache zu generieren (so. truncated hypothesis | k.A.  |



|                      |     |                                                              |            |                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                              |
|----------------------|-----|--------------------------------------------------------------|------------|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
|                      |     |                                                              |            |                                                                  | generation).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                              |
| BEDARD/BIGGS (1991b) | 176 | 79 Manager<br>Ø = 7,2 J.<br><br>97 Prüfungsleiter<br>Ø = 3,5 | Experiment | Chi-Quadrat-Test                                                 | Prüfer ohne Mandantenerklärung waren nicht signifikant erfolgreicher als Prüfer mit inkorrekt er Mandantenerklärung.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 0,793                                                                                                                        |
| BEDARD/BIGGS (1991b) | 176 | 79 Manager<br>Ø = 7,2 J.<br><br>97 Prüfungsleiter<br>Ø = 3,5 | Experiment | Chi-Quadrat-Test<br><br>Chi-Quadrat-Test<br><br>Chi-Quadrat-Test | Allgemeine Prüfungserfahrung hat keinen signifi-<br>kanten Einfluss auf die Erfolgsquote der Prüfer.<br>Domänenspezifische Erfahrung hat einen signifi-<br>kanten Einfluss auf die Erfolgsquote der Prüfer.<br>Domänenspezifische Erfahrung hat nur dann ein-<br>nen signifikanten Einfluss auf die Erfolgsquote<br>der Prüfer, wenn die Mandantenerklärung korrekt<br>ist, nicht hingegen, wenn die Mandantenerklä-<br>rung inkorrekt ist, bzw. wenn keine Mandanten-<br>erklärung mitgeteilt wird.                                                                                                | 0,708<br><br>0,022 <sup>1</sup> (0,056 <sup>2</sup> )<br><br>0,002 (0,025)<br>ggü.<br>0,610 (0,467)<br>bzw.<br>0,817 (0,908) |
| GREEN (2004b)        | 58  | k.A.<br>Ø = 4,0 J.                                           | Experiment | Varianzanalyse<br><br><br><br><br><br><br><br>Chi-Quadrat-Test   | Branchenspezialisten, die eine inkorrekte Nicht-<br>Fehlerhypothese des Mandanten erhielten, gene-<br>rierten anschließend weder signifikant mehr<br>Nicht-Fehlerhypothesen, noch signifikant weniger<br>Fehlerhypothesen, als Nicht-Spezialisten, die<br>ebenfalls eine inkorrekte Nicht-Fehlerhypothese<br>des Mandanten erhielten.<br>Branchenspezialisten, die eine inkorrekte Nicht-<br>Fehlerhypothese des Mandanten erhielten, gene-<br>rierten anschließend nicht signifikant häufiger die<br>korrekte Abweichungsursache, als Nicht-<br>Spezialisten, die ebenfalls eine inkorrekte Nicht- | 0,274 bzw.<br>0,470<br><br><br><br><br><br><br>0,724                                                                         |

<sup>1</sup> Hierbei wurde als Maß für domänenspezifische Erfahrung die Anzahl der betreuten Mandanten mit wesentlichem Vorratsvermögen herangezogen.

<sup>2</sup> Hierbei wurde als Maß für domänenspezifische Erfahrung die Anzahl der betreuten Mandanten aus der Industriebranche herangezogen.

|                         |     |                                                                    |            |                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                             |
|-------------------------|-----|--------------------------------------------------------------------|------------|----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
|                         |     |                                                                    |            | Chi-Quadrat-Test                       | Fehlerhypothese des Mandanten erhielten. Branchenspezialisten, die die korrekte Hypothese generierten, waren signifikant häufiger dazu in der Lage, die korrekte Abweichungsursache zu identifizieren, als Nicht-Spezialisten, die die korrekte Hypothese generierten.                                                                                                                                                                                            | 0,019                                                       |
| LIBBY (1985)            | 37  | Manager<br>Ø = 5 J.                                                | Experiment | Varianzanalyse                         | Die Vererbung einer Fehlerhypothese durch einen Vorgesetzten führt nicht zu einer vermehrten anschließenden Generierung von weiteren Fehlerhypothesen.                                                                                                                                                                                                                                                                                                            | ≤ 0,9294                                                    |
| LIBBY/FREDERICK (1990)  | 126 | 61 Manager<br>Median = 5 J.<br><br>70 Assistenten<br>Median = 1 J. | Experiment | Varianzanalyse<br><br>Varianzanalyse   | Prüfer, die eine atypische Fehlerhypothese von ihrem Vorgesetzten erhielten, generierten anschließend signifikant mehr Fehlerhypothesen des gleichen Transaktionskreises, als Prüfer, denen keine Erklärung vererbt wurde.<br><br>Prüfer, die eine typische Fehlerhypothese von ihrem Vorgesetzten erhielten, generierten anschließend nicht signifikant mehr Fehlerhypothesen des gleichen Transaktionskreises, als Prüfer, denen keine Erklärung vererbt wurde. | 0,09 <sup>1</sup> /0,076 <sup>2</sup><br><br>> 0,1          |
| CHURCH/SCHNEIDER (1993) | 96  | k.A.<br>Ø = 2 J.                                                   | Experiment | Varianzanalyse<br><br>Chi-Quadrat-Test | Prüfer, die eine Fehlerhypothese von einem Vorgesetzten erhalten, generieren anschließend signifikant weniger Fehlerhypothesen des gleichen Transaktionskreises, als Prüfer, denen keine Fehlerhypothese vererbt wurde.<br><br>Prüfer, die eine Fehlerhypothese von einem Vorgesetzten erhalten, generieren anschließend signifikant häufiger Fehlerhypothesen eines an-                                                                                          | 0,0336<br><br>0,0002 <sup>1</sup><br>(0,0011 <sup>2</sup> ) |

<sup>1</sup> Signifikanzniveau bei Managern.

<sup>2</sup> Signifikanzniveau bei Prüfungsassistenten.

|                             |     |                                                                                          |            |                  |                                                                                                                                                                                                                                                         |                                          |
|-----------------------------|-----|------------------------------------------------------------------------------------------|------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
|                             |     |                                                                                          |            |                  | deren Transaktionskreises, als Prüfer, denen keine Fehlerhypothese vererbt wurde.                                                                                                                                                                       |                                          |
| MUELLER/ANDERSON (2002)     | 65  | k.A.<br>Ø = 5,5 J.                                                                       | Experiment | Kovarianzanalyse | Prüfer, die bei der Auswahl von potenziellen Abweichungsursachen aus einer vorgegebenen Liste eine Inklusionsstrategie anwenden, nehmen eine signifikant stärkere Reduzierung der Ausgangsliste vor, als Prüfer, die eine Exklusionsstrategie anwenden. | < 0,001                                  |
| ISMAIL/TROTMAN (1995)       | 152 | 42 Manager<br>Ø = 6,3<br><br>34 Prüfungsleiter<br>Ø = 3,4<br><br>38 Prüfer<br>Ø = 1,8 J. | Experiment | t-Test           | Der Review-Prozess führt zu einer signifikant höheren Anzahl plausibler Fehlerhypothesen.                                                                                                                                                               | 0,000 <sup>3</sup> /0,006 <sup>4</sup> / |
|                             |     |                                                                                          |            | Varianzanalyse   | Der Review-Prozess mit Diskussion führt zu einer signifikant höheren Anzahl plausibler Hypothesen als der Review-Prozess ohne Diskussion.                                                                                                               | 0,000 <sup>5</sup> /0,005 <sup>6</sup>   |
|                             |     |                                                                                          |            | Varianzanalyse   | Der Review-Prozess mit Diskussion führte nicht signifikant mehr plausible Hypothesen hinzu als der Review-Prozess ohne Diskussion.                                                                                                                      | < 0,10                                   |
|                             |     |                                                                                          |            | Varianzanalyse   | Der Review-Prozess mit Diskussion verwarf signifikant weniger plausible Hypothesen als der Review-Prozess ohne Diskussion.                                                                                                                              | 0,593                                    |
| BEDARD/BIGGS/MARONEY (1998) | 35  | Prüfungsleiter<br>k.A.                                                                   | Experiment | Chi-Quadrat-Test | Prüfergruppen generieren signifikant häufiger die richtige Abweichungshypothese als einzelne Prüfer.                                                                                                                                                    | 0,004                                    |
|                             |     |                                                                                          |            | Chi-Quadrat-Test | Prüfergruppen zogen signifikant häufiger die richtigen Schlussfolgerungen und generierten mehr                                                                                                                                                          | 0,013                                    |
|                             |     |                                                                                          |            |                  |                                                                                                                                                                                                                                                         | 0,018                                    |

<sup>1</sup> Signifikanzniveau bei Prüfern, denen eine Fehlerhypothese aus dem Transaktionskreis Forderungen/Umsätze vererbt wurde.

<sup>2</sup> Signifikanzniveau bei Prüfern, denen eine Fehlerhypothese aus dem Transaktionskreis Verbindlichkeiten/Einkäufe vererbt wurde.

<sup>3</sup> Signifikanzniveau bei Review durch einen Prüfungsleiter mit Diskussion.

<sup>4</sup> Signifikanzniveau bei Review durch einen Prüfungsleiter ohne Diskussion.

<sup>5</sup> Signifikanzniveau bei Review durch einen Manager mit Diskussion.

<sup>6</sup> Signifikanzniveau bei Review durch einen Manager ohne Diskussion.

|  |  |  |  |                    |                                                                                                                                                                                                      |      |
|--|--|--|--|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
|  |  |  |  | Chi-Quadrat-Test   | plausible Hypothesen als einzelne Prüfer.<br>Gruppen und Einzelprüfer korrigierten identisch häufig die begangenen Fehler.<br>Gruppen waren nicht signifikant häufiger erfolgreich als Einzelprüfer. | k.A. |
|  |  |  |  | Verbale Protokolle | Nicht erfolgreiche Gruppen wiesen die höchste Anzahl von Unterbrechungen auf. Erfolgreiche Gruppen wiesen keine Unterbrechungen auf.                                                                 | k.A. |
|  |  |  |  | Verbale Protokolle | Dominanz einzelner Gruppenmitglieder wirkt sich positiv auf die Hypothesengenerierung und -selektion aus.                                                                                            | k.A. |

Tabelle 3: Ergebnisse der empirischen Prüfungsforschung zur Phase der Hypothesengenerierung

### 3.5.4 Studien zur Phase der Informationssuche

#### 3.5.4.1 Vorbemerkungen

Im Anschluss an die Hypothesengenerierung wird sich der Prüfer auf die Suche nach weiteren Informationen begeben, um Aufschluss über die Gültigkeit bzw. Ungültigkeit seiner Hypothesen zu erhalten. Im Kontext der Jahresabschlussprüfung geschieht dieses regelmäßig durch die Erlangung von Prüfungsnachweisen.

Obwohl sich ein großer Teil der Prüfungsforschung zunächst überwiegend auf die Hypothesengenerierung konzentrierte und dieser Phase eine überragende Bedeutung beigemessen wurde, setzte sich allmählich die Meinung durch, dass auch der darauf folgenden Phase der Informationssuche eine höhere Aufmerksamkeit zu widmen sei. Diese Erkenntnis beruhte im Wesentlichen auf drei Beobachtungen.<sup>1</sup> Erstens, falls der Prüfer eine inkorrekte Hypothese abrufte oder „vererbt“, können konfirmatorische Prozesse<sup>2</sup> zu einer unzweckmäßigen voreingenommenen Informationssuche führen. Zweitens, die Generierung einer richtigen Hypothese bedeutet nicht zwangsläufig, dass die analytische Prüfungshandlung auch erfolgreich abgeschlossen wird. Vielmehr kann eine unzweckmäßige Informationssuche zu falschen Schlussfolgerungen führen.<sup>3</sup> Drittens, auch wenn der Prüfer zunächst nicht die richtige Fehlerhypothese generiert, ermöglicht eine zweckmäßige Informationssuche, dass der Prüfer seine anfängliche Schwäche kompensiert und durch die Erlangung der richtigen Informationen zu einem späteren Zeitpunkt die richtige Abweichungsursache abrufte.<sup>4</sup>

Die Informationssuche kann dabei generell in drei unterschiedliche Suchstrategien differenziert werden. In Betracht kommen dabei die positive, die negative sowie die ausgewogene (balanced) Suchstrategie. Die *positive Suchstrategie* ist überwiegend auf eine Bestätigung der generierten Hypothese aus und fokussiert deshalb Prüfungstests, die Hinweise auf die Richtigkeit der Hypothese erwarten lassen (sog. *confirming evidence*).<sup>5</sup> Die Verwendung der positiven Suchstrategie geht i.d.R. mit der Existenz konfirmatorischer Prozesse (sog. *confirmation bias*) einher.<sup>6</sup> Hierbei werden Hinweise, die auf die Gültigkeit der Hypothese hindeuten, überbetont bzw. Informationen, die auf die In-

---

<sup>1</sup> Vgl. im Folgenden ASARE/WRIGHT (2001), S. 208.

<sup>2</sup> Hinweise auf sog. Bestätigungseffekte finden sich auch in Kap. 3.5.2.3.

<sup>3</sup> Vgl. hierzu auch die Ergebnisse in Kap. 3.5.3.2.

<sup>4</sup> Auch hierzu soll auf die Ergebnisse in Kap. 3.5.3.2 verwiesen werden.

<sup>5</sup> Vgl. KLAYMAN/HA (1987), S. 211-228. Häufig wird diese Strategie als *truncated search strategy* bezeichnet. Vgl. etwa KOONCE (1993), S. 65 oder HO (1994), S. 211.

<sup>6</sup> Ähnlich argumentieren PETERSON/WONG-ON-WING (2000), S. 257 m.w.N.

korrektheit der eigenen Hypothese hindeuten, entweder „heruntergespielt“ oder falsch interpretiert (sog. „explaining away“ disconfirming evidence). Die *negative Suchstrategie* ist hingegen dadurch gekennzeichnet, dass Prüfer vornehmlich Prüfungstests durchführen, von denen sie Hinweise erwarten, welche die eigene Hypothese widerlegen bzw. auf die Gültigkeit einer anderen Hypothese hindeuten (sog. disconfirming evidence).<sup>1</sup> Eine *ausgewogene Suchstrategie* favorisiert hingegen weder confirming evidence noch disconfirming evidence. Sie versucht vielmehr Hinweise zu erhalten, die sich auf eine Vielzahl von potenziellen Abweichungsgründen beziehen.<sup>2</sup>

In der Prüfungsforschung wurde der Untersuchung der positiven Teststrategie bzw. von konfirmatorischen Prozessen die größte Bedeutung beigemessen. Daher soll im Folgenden eine ausführliche Darstellung der Studien vorgenommen werden, die sich mit den potenziellen Verzerrungen im Rahmen der Informationssuche beschäftigen. Anschließend sollen weitere Studien dargestellt werden, die untersuchen, welche Suchstrategie besonders zweckmäßig erscheint.

### **3.5.4.2 Verzerrungen im Rahmen der Informationssuche**

#### **3.5.4.2.1 Konfirmatorische Prozesse und positive Suchstrategie**

Haben Prüfer eine Abweichungshypothese intern oder extern generiert, so besteht tendenziell die Gefahr, dass sie ihre Informationssuche von vornherein auf die Bestätigung dieser Hypothese ausrichten. In der psychologischen Forschung wurden diese konfirmatorischen Prozesse bereits mehrfach nachgewiesen.<sup>3</sup>

In der Prüfungsforschung wurde die Suchstrategie erstmalig von KAPLAN/RECKERS untersucht. Die Autoren ließen 71 Prüfer mit einer durchschnittlichen Prüfungserfahrung von acht Jahren einen Fall bearbeiten, bei dem ein nicht eingebuchter Wareneingang zu einer unerwarteten wesentlichen Abweichung in der Bruttogewinnmarge und der Umschlagshäufigkeit der Vorräte führte.<sup>4</sup> Die Prüfer wurden zunächst gefragt, für wie hoch sie die Wahrscheinlichkeit einschätzen, dass die Diskrepanz auf einen Fehler bzw. Nicht-Fehler zurückzuführen ist (sog. initial beliefs). Anschließend wurde ihnen eine

---

<sup>1</sup> Vgl. KLAYMAN/HO (1989), S. 596-604.

<sup>2</sup> Vgl. ASARE/WRIGHT (2003), S. 235.

<sup>3</sup> Vgl. KANOUSE (1972), S. 121-135; MAJOR (1980), S. 1010-1023; SHAKLEE/FISCHHOFF (1982), S. 520-530; TAPLIN (1975), S. 85-96.

<sup>4</sup> Vgl. KAPLAN/RECKERS (1989), S. 543-544.

Liste mit 20 Prüfungstests zur Verfügung gestellt, von denen sie die sechs zur Identifikation der Abweichungsursache relevantesten Tests in eine Rangordnung bringen sollten.<sup>1</sup>

Im Ergebnis stellten KAPLAN/RECKERS wesentliche Differenzen zwischen der Suchstrategie der unerfahrenen und erfahrenen Prüfer fest. Unerfahrene Prüfer (d.h. Prüfer mit max. fünf Jahren Prüfungserfahrung) konzentrierten sich überwiegend auf Tests, die auf die Bestätigung ihrer ursprünglichen Überzeugung (initial beliefs) abzielten.<sup>2</sup> Hielten es diese Prüfer für wahrscheinlicher, dass die festgestellte Abweichung durch einen Fehler (Nicht-Fehler) verursacht wurde, so wählten sie signifikant mehr Tests aus, die mit der Überprüfung eines Fehlers (Nicht-Fehlers) zusammenhingen. Insofern offenbarten unerfahrene Prüfer eine positive Suchstrategie bzw. konfirmatorische Prozesse.<sup>3</sup> Keinen signifikanten Einfluss zeigten hingegen die anfänglichen Überzeugungen der erfahrenen Prüfer (d.h. Prüfer mit sechs oder mehr Jahren Prüfungserfahrung). Diese Prüfergruppe praktizierte unabhängig davon, ob sie von einer Fehler- oder Nicht-Fehlerursache ausgingen, eine ausgewogene Suchstrategie.<sup>4</sup>

In der Folgezeit wurden zahlreiche Studien im Prüfungsbereich durchgeführt, die weitere Indizien für die Existenz von konfirmatorischen Prozessen erbrachten. BEDARD/BIGGS etwa stellten fest, dass 38,1% der Probanden, ihre selbstgenerierte Hypothese nicht verwerfen konnten, obwohl bereits diverse Hinweise vorlagen, die die Gültigkeit ihrer Hypothese widerlegten.<sup>5</sup> Ähnliche Befunde lieferten HEIMAN-HOFFMAN/MOSER/JOSEPH.<sup>6</sup>

Konfirmatorische Prozesse konnten allerdings auch beim Erhalt einer inkorrekten Mandantenerklärung nachgewiesen werden. ANDERSON/KOONCE etwa stellten fest, dass sich Prüfer trotz des Vorliegens gegenteiliger Beweise weiterhin bemüht zeigten, Argumente für die Aufrechterhaltung der Hypothese zu erbringen, indem entsprechende Hinweise überbetont oder fehlinterpretiert wurden (sog. „explaining away“ disconfirming eviden-

---

<sup>1</sup> Vgl. KAPLAN/RECKERS (1989), S. 544-545.

<sup>2</sup> Vgl. KAPLAN/RECKERS (1989), S. 547.

<sup>3</sup> Vgl. KAPLAN/RECKERS (1989), S. 549.

<sup>4</sup> Vgl. KAPLAN/RECKERS (1989), S. 547.

<sup>5</sup> Vgl. BEDARD/BIGGS (1991a), S. 640. Für eine ausführliche Darstellung der Studie soll auf Kap. 3.5.2.1 verwiesen werden.

<sup>6</sup> Vgl. HEIMAN-HOFFMAN/MOSER/JOSEPH (1995), S. 776. Eine ausführliche Darstellung ihrer Untersuchung findet sich in Kap. 3.5.3.3.1.4.

ce).<sup>1</sup> Einen derartigen Bestätigungseffekt konnten auch BIERSTAKER/BEDARD/BIGGS nachweisen.<sup>2</sup> Darüber hinaus stellten sie auch fest, dass einige Probanden trotz des Vorliegens von Hinweisen, die nicht mit der Richtigkeit der Mandantenerklärung vereinbart werden konnten, immer und immer die Mandantenerklärung evaluierten (sog. wheel spinning).<sup>3</sup>

Dass eine positive Suchstrategie nicht zwangsläufig negative Auswirkungen mit sich bringen muss, versuchten PETERSON/WONG-ON-WING nachzuweisen. Vielmehr sei es davon abhängig, welche Aufgabenstruktur dem Sachverhalt zugrunde liegt. Spiegelt die richtige Hypothese im Wesentlichen den tatsächlichen Sachverhalt wieder, so kann nicht bestritten werden, dass die positive Suchstrategie ein geeignetes Mittel darstellt, um die Korrektheit der Hypothese festzustellen.<sup>4</sup> Um diesen Zusammenhang auch für andere Aufgabenstrukturen empirisch zu belegen, differenzierten die Autoren in ihrer Studie zwischen einer eingebetteten und einer umschlossenen Aufgabenstruktur.<sup>5</sup> Eine *eingebettete Aufgabenstruktur* liegt vor, wenn die generierte Hypothese einen Teilbereich des tatsächlichen Abweichungsgrunds darstellt (etwa, wenn der Prüfer einen Fehler im Transaktionskreis Forderungen & Umsätze vermutet, die festgestellte Abweichung jedoch tatsächlich sowohl durch einen Fehler in den Forderungen & Umsätzen als auch durch einen Fehler im Bereich der Verbindlichkeiten & Beschaffungen verursacht wurde). Hier kann unterstellt werden, dass eine positive Suchstrategie niemals zu einer Falsifizierung der eigenen teilweise falschen Hypothese führen kann.

Von einer *umschlossenen Aufgabenstruktur* ist hingegen auszugehen, wenn die tatsächliche Abweichungsursache einen Teilbereich der generierten Hypothese darstellt (etwa, wenn der Prüfer davon ausgeht, dass die Abweichung sowohl durch einen Fehler in den Forderungen & Umsätzen als auch durch einen Fehler im Bereich der Verbindlichkeiten & Beschaffungen hervorgerufen wurde, die tatsächliche Abweichung jedoch durch einen Fehler in nur einem der beiden Bereiche hervorgerufen wurde). In diesem Fall wäre davon auszugehen, dass eine positive Suchstrategie es dem Prüfer ermöglicht, die teilweise Inkorrektheit der eigenen Hypothese zu erkennen.

---

<sup>1</sup> Vgl. ANDERSON/KOONCE (1995), S. 130. Da der Schwerpunkt ihrer Studie in der Phase der Hypothesenbewertung zu sehen ist, soll an dieser Stelle von einer vertieften Darstellung abgesehen und in Kap. 3.5.5.3.1.2.1.2 nachgeholt werden.

<sup>2</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 32. Eine ausführliche Darstellung der Studie findet sich unter Kap. 3.5.2.3 i.V.m. 3.5.2.1.

<sup>3</sup> Vgl. BIERSTAKER/BEDARD/BIGGS (1999), S. 32.

<sup>4</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 274-275 m.w.N.

<sup>5</sup> Vgl. im Folgenden PETERSON/WONG-ON-WING (2000), S. 261-264.



In diesem Zusammenhang ließen PETERSON/WONG-ON-WING insgesamt 93 Prüfer entweder einen Fall mit einer eingebetteten Aufgabenstruktur oder mit einer umschlossenen Aufgabenstruktur bearbeiten.<sup>1</sup> Dabei stellten die Autoren fest, dass Prüfer, die eine eingebettete Version bearbeiteten und zudem ein positives Suchverhalten offenbarten, marginal weniger effizient waren. Bei den Prüfern, die der umschlossenen Aufgabenstruktur zugeordnet wurden und die zudem einer positiven Suchstrategie nachgingen, ließen die Ergebnisse erkennen, dass sie signifikant effizienter waren.<sup>2</sup> Insofern konnte bestätigt werden, dass eine positive Suchstrategie nicht grundsätzlich zu unerwünschten Auswirkungen führt und unter bestimmten Bedingungen vorteilhaft erscheint.

#### **3.5.4.2.2 Commitment als wesentlicher Einflussfaktor der Verzerrungen**

CHURCH bemühte sich als Erster, konfirmatorische Prozesse sowie die Verwendung der positiven Suchstrategie theoretisch zu fundieren. Dabei stellte sich heraus, dass diese Verzerrungen in erster Linie von der Bindung (commitment) des Prüfers an die generierte Hypothese abhängen, welche wiederum maßgeblich von motivationalen Aspekten beeinflusst wird.<sup>3</sup>

Aus motivationaler Sicht streben Individuen nach der *Beibehaltung der kognitiven Konsistenz*. Da inkonsistente Kognitionen erregend und psychologisch unangenehm sind, neigen Individuen dazu, gegenteilige Beweise herunterzuspielen. Ein weiterer motivationaler Aspekt ist in der *Suche nach Annerkennung von Anderen* zu sehen. Da Individuen ihr Ansehen in der Umwelt nicht gefährden wollen, tendieren sie dazu, ihre Meinung möglichst lang zu verteidigen.

Um diesen Zusammenhang empirisch zu belegen, ließ CHURCH 79 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 2,5 Jahren einen Fall bearbeiten, der eine unerwartete wesentliche Abweichung in der Bruttogewinnmarge umfasste. Den Prüfern wurde mitgeteilt, dass im Rahmen der Vorprüfungen diverse Mängel im internen Kontrollsystem festgestellt werden konnten und die identifizierte Diskrepanz entweder auf einen Fehler im Bereich der Forderungen & Umsätze oder im Bereich der Verbindlich-

---

<sup>1</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 267.

<sup>2</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 272.

<sup>3</sup> Vgl. im Folgenden CHURCH (1990), S. 83-91.

keiten & Beschaffungen zurückzuführen sei. Schließlich wurden sie aufgefordert, eine Fehlerhypothese aufzustellen.<sup>1</sup>

Anschließend wurde ein Teil der Prüfer gebeten, die Argumente für ihre geäußerte Annahme als Grundlage für eine folgende Diskussion mit einem Vorgesetzten zu dokumentieren. Da zudem der Name des Prüfers zu vermerken war, ging CHURCH davon aus, dass diese Prüfergruppe eine starke Bindung zu ihrer Abweichungshypothese aufwies. Der andere Teil der Probanden musste seine Argumente hingegen nicht dokumentieren. Zudem wurde weder eine anschließende Diskussion mit einem Vorgesetzten, noch der Name des Probanden verlangt. CHURCH ging davon aus, dass diese Prüfergruppe keine starke Bindung zu ihrer Fehlerhypothese offenbarte.<sup>2</sup>

Daraufhin erhielten beide Prüfergruppen zehn weitere Testergebnisse zu den internen Kontrollen des Unternehmens. Da die Hinweise sowohl auf Probleme bei den Forderungen & Umsätzen als auch auf Probleme bei den Verbindlichkeiten & Beschaffungen hindeuteten, ließen diese zusätzlichen Informationen keine eindeutigen Schlussfolgerungen hinsichtlich des tatsächlichen Abweichungsgrunds zu. Nachdem die Prüfer die Informationen zur Kenntnis genommen hatten, wurden sie schließlich aufgefordert, 20 weitere Prüfungsstunden auf die beiden relevanten Transaktionskreise zu verteilen. Abschließend wurden alle Prüfer gefragt, welche Testergebnisse sie als besonders relevant erachteten, um die tatsächliche Abweichungsursache zu identifizieren.<sup>3</sup>

Im Ergebnis stellte CHURCH einen signifikanten Zusammenhang zwischen commitment und konfirmatorischen Prozessen fest. Prüfer mit einer starken Bindung verteilten im Schnitt 11,2 Stunden auf den ausgewählten Transaktionskreis und Prüfer ohne Bindung hingegen 10,02 Stunden.<sup>4</sup> Damit ist die positive Suchstrategie der Prüfer mit Bindung stärker als bei Prüfern ohne Bindung. Darüber hinaus ordneten Prüfer mit Bindung diejenigen Informationen, welche auf Probleme in dem von ihnen ausgewählten Bereich hinwiesen (sog. confirming evidence), eine signifikant höhere Bedeutung zu als Prüfer ohne Bindung.<sup>1</sup> Dieses stützt die Annahme, dass Prüfer mit Bindung zu konfirmatorischen Prozessen neigen. Allerdings waren zwischen Prüfern mit und Prüfern ohne Bindung keine signifikanten Unterschiede hinsichtlich der zugeordneten Bedeutung von Testergebnissen zu verzeichnen, die auf Probleme in dem nicht ausgewählten Transak-

---

<sup>1</sup> Vgl. CHURCH (1991), S. 519-520.

<sup>2</sup> Vgl. CHURCH (1991), S. 520.

<sup>3</sup> Vgl. CHURCH (1991), S. 521.

<sup>4</sup> Vgl. CHURCH (1991), S. 523-524.

tionskreis hinwiesen, so dass CHURCH davon ausging, dass sog. disconfirming evidence von Prüfern mit Bindung ähnlich verwertet wurden wie von Prüfern ohne Bindung.<sup>2</sup>

PETERSON/WONG-ON-WING merkten diesbezüglich an, dass das hinsichtlich der Interpretation von disconfirming evidence unerwartete Ergebnis von CHURCH möglicherweise darauf zurückzuführen ist, dass die Klassifizierung in confirming bzw. disconfirming evidence in Relation zur anfänglichen Hypothesengenerierung betrachtet wurde.<sup>3</sup> Da Prüfer jedoch auch eine Hypothese wechseln bzw. nachträglich generieren können,<sup>4</sup> sollte eine Differenzierung der Suchstrategie in Bezug zur aktuell generierten Hypothese stattfinden.<sup>5</sup>

Um ihre Vermutung zu verifizieren, ließen die Autoren 93 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,3 Jahren eine Aufgabe bearbeiten, bei der ein Fehler im Transaktionskreis der Forderungen & Umsätze und/oder der Verbindlichkeiten & Beschaffungen zu einer wesentlichen Abweichung in der Bruttogewinnmarge führte.<sup>6</sup> Im Unterschied zu CHURCH wurden die Probanden nach der Kenntnisnahme der Testergebnisse erneut aufgefordert, eine Abweichungshypothese zu benennen.<sup>7</sup> Dieses ermöglichte es den Autoren, die Suchstrategie nach der aktuell generierten Hypothese zu klassifizieren.

Im Ergebnis ließen die Befunde von PETERSON/WONG-ON-WING erkennen, dass eine Zuteilung der Suchstrategie nach der aktuellen Hypothese im Vergleich zu einer Zuteilung nach der anfänglichen Hypothese zu einer signifikant höheren Anzahl von durchgeführten Tests führte, die dem ausgewählten Transaktionskreis zuzuordnen sind (sog. positive Suchstrategie; 73,7% im Vgl. zu 59,2%) und zu einer signifikant niedrigeren Anzahl von Tests, die einem anderen Transaktionskreis anzurechnen sind (sog. negative Suchstrategie; 11,8% im Vgl. zu 29,0%).<sup>8</sup> Diese Ergebnisse zeigen, dass eine wie von CHURCH vorgenommene Zuteilung nach der anfänglichen Hypothese zu weniger eindeutigen Ergebnissen führt als eine Zuteilung nach der aktuellen Hypothese. Da Letzte-

---

<sup>1</sup> Vgl. CHURCH (1991), S. 526-527.

<sup>2</sup> Vgl. CHURCH (1991), S. 528-529.

<sup>3</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 258-259.

<sup>4</sup> Zum Aspekt des Hypothesenwechsels sei auf Kap. 3.5.2.3 sowie 3.5.3.3.1.4 verwiesen. Dass Prüfer Hypothesen auch nachträglich, also im Rahmen der weiteren prüferischen Vorgehensweise generieren, wurde ausführlich in Kap. 3.5.3.2 beschrieben.

<sup>5</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 274.

<sup>6</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 266. Da der weitere Versuchsaufbau im Wesentlichen mit der Aufgabe von CHURCH (1991) verglichen werden kann, soll an dieser Stelle von weiteren Einzelheiten abgesehen werden.

<sup>7</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 267.

<sup>8</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 270.

re jedoch die zweckmäßigere Vorgehensweise darstellt, konstatierten die Autoren, dass Prüfer tendenziell zu einer positiven Suchstrategie neigen, welche auf die Bestätigung ihre Hypothese ausgerichtet ist.<sup>1</sup>

#### **3.5.4.3    Eigenschaften der Informationssuche und ihr Einfluss auf den Erfolg analytischer Prüfungshandlungen**

Eine Studie, die den Einfluss von unterschiedlichen Sucheigenschaften auf den Erfolg von analytischen Prüfungshandlungen untersuchte, führten ASARE/TROMPETER/WRIGHT durch. Als Sucheigenschaften wurden dabei der Umfang, die Breite und die Tiefe der Informationssuche herangezogen.<sup>2</sup> Der *Umfang* bezieht sich dabei auf die Anzahl der durchgeführten Prüfungstests. Die *Breite* untersucht die Anzahl der getesteten Hypothesen. Die *Tiefe* beschäftigt sich schließlich mit der Anzahl der Tests pro Hypothese.

Im Kontext der zuvor diskutierten Suchstrategien kann davon ausgegangen werden, dass eine breit angelegte Informationssuche in einem unmittelbaren Zusammenhang zu der ausgewogenen Suchstrategie steht. Bei einer positiven bzw. negativen Suchstrategie ist hingegen anzunehmen, dass sie einer tiefen Informationssuche gleichgestellt werden kann.

In diesem Zusammenhang ließen ASARE/TROMPETER/WRIGHT 91 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 2,9 Jahren an einer Studie teilnehmen, bei der eine fehlerhafte Gemeinkostenallokation zu einer unerwarteten wesentlichen Abweichung in der Bruttogewinnmarge führte. Die Prüfer wurden aufgefordert, eine von fünf präsentierten Hypothesen zu wählen, die ihrer Ansicht nach zur festgestellten Diskrepanz führte. Anschließend konnten sie jeweils einen Test von einer Liste mit 15 möglichen Prüfungstests auswählen und die entsprechenden Testergebnisse einsehen. Dieser Vorgang konnte solange wiederholt werden, bis sich die Prüfer in der Lage fühlten, eine endgültige Abweichungshypothese zu benennen.<sup>3</sup>

Im Ergebnis konnten die Autoren feststellen, dass ausschließlich die Breite der Informationssuche einen signifikanten Einfluss auf den Erfolg der analytischen Prüfungshandlung hatte. Bei der Tiefe der Suche konnte hingegen kein signifikanter Zusammenhang

---

<sup>1</sup> Vgl. PETERSON/WONG-ON-WING (2000), S. 274.

<sup>2</sup> Vgl. im Folgenden ASARE/TROMPETER/WRIGHT (2000), S. 544, 546-548.

<sup>3</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 549-551.

zum Erfolg nachgewiesen werden.<sup>1</sup> Ein vergleichbares Ergebnis liefern zudem ASARE/WRIGHT, die feststellten, dass erfolgreiche Prüfer im Schnitt signifikant mehr Hypothesen testeten als nicht erfolgreiche Prüfer (3,91 im Vgl. zu 3,24).<sup>2</sup> Darüber hinaus konnten GREEN/TROTMAN in einer anderen Studie erkennen, dass sich die Tiefensuche signifikant negativ auf den Erfolg der teilnehmenden Prüfer auswirkte.<sup>3</sup> Je höher die Anzahl der Tests pro Hypothese ausfiel, desto weniger waren die Prüfer dazu in der Lage, die Aufgabe erfolgreich zu beenden.<sup>4</sup> Da eine vertiefte Suche im Wesentlichen durch umfangreiches (positives) Testen der in Betracht gezogenen Hypothese beeinflusst wurde, kann davon ausgegangen werden, dass die jeweiligen Prüfer einer positiven Teststrategie folgten.

Die Autoren nahmen ihre empirischen Befunde zum Anlass, um auf die Bedeutung einer Breitensuche bzw. einer ausgewogenen Teststrategie für den Erfolg von analytischen Prüfungshandlungen hinzuweisen. Darüber hinaus legen die Ergebnisse nahe, dass eine Tiefensuche bzw. positive Teststrategie negative Effekte für den Erfolg der Prüfer nach sich zieht. Die Prüfungspraxis sollte folglich darauf bedacht sein, ihre Prüfer zu einer breit angelegten Informationssuche zu bewegen.

Um diese Befunde weiter zu manifestieren, führten ASARE/WRIGHT ebenfalls eine Studie durch, die sich mit dem Einfluss des Suchverhaltens auf den Erfolg bei analytischen Prüfungshandlungen befasste.<sup>1</sup> Dabei sollten die 102 teilnehmenden Prüfer mit einer durchschnittlichen Prüfungserfahrung von 2,05 Jahren ebenfalls die Ursache für eine durch eine fehlerhafte Gemeinkostenallokation verursachte Abweichung in der Bruttogewinnmarge identifizieren. Im Vergleich zu ASARE/TROMPETER/WRIGHT erweiterten ASARE/WRIGHT ihre Untersuchung um eine Variation der Hypothesengenerierung. Eine Prüfergruppe wurde aufgefordert, fünf potenzielle Abweichungsursachen eigenständig zu generieren. Zwei weiteren Prüfergruppen wurden hingegen fünf potenzielle Hypothesen zur Verfügung gestellt, wobei der Hypothesensatz der einen Gruppe den tatsächlichen Abweichungsgrund enthielt, der Hypothesensatz der anderen Gruppe hingegen nicht. Daraufhin wurden die Probanden aufgefordert, die Hypothese zu benennen, von der sie annehmen, dass sie für die Diskrepanz verantwortlich ist. Im Anschluss an die

---

<sup>1</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 555-556.

<sup>2</sup> Vgl. ASARE/WRIGHT (2003), S. 246. Eine ausführliche Darstellung der Vorgehensweise erfolgt weiter unten im vorliegenden Abschnitt.

<sup>3</sup> Für eine ausführliche Darstellung der Ergebnisse sei auf Kap. 3.5.3.2 verwiesen.

<sup>4</sup> Vgl. GREEN/TROTMAN (2003), S. 228-229. Allerdings stellten die Autoren entgegen der Ergebnisse von ASARE/TROMPETER/WRIGHT (2000) sowie ASARE/WRIGHT (2003) auch fest, dass die Breite der Suche keinen wesentlichen Einfluss auf den Erfolg der Prüfer mit sich brachte.

Hypothesengenerierung erfolgte die Phase der Informationssuche. Hierbei wurde ein Teil der Prüfer aufgefordert, eine eigenständige Suche vorzunehmen. Dieser Prüfergruppe war es erlaubt, sechs Prüfungstests aus einer Liste mit 21 möglichen Tests sequentiell auszuwählen und einzusehen. Dem anderen Teil der Prüfer wurde hingegen eine Liste mit den Ergebnissen von sechs Prüfungstests vorgegeben. Da sich die Testergebnisse auf eine breite Anzahl von möglichen Abweichungsursachen bezogen, betrachten die Autoren dieses als das Resultat einer ausgewogenen Suchstrategie. Nachdem die Prüfer alle Testergebnisse zur Kenntnis genommen hatten, wurden sie erneut aufgefordert, eine Hypothese zu selektieren, die sie für den Grund der festgestellten Abweichung halten.

Im Ergebnis stellten ASARE/WRIGHT fest, dass Prüfer mit dem richtigen Hypothesensatz (d.h. der extern zur Verfügung gestellte Hypothesensatz umfasst die tatsächliche Abweichungsursache) und ausgewogenen Testergebnissen marginal erfolgreicher waren als Prüfer mit dem richtigen Hypothesensatz und eigenständiger Hypothesensuche (71 % im Vgl. zu 47%).<sup>2</sup> Der Grund für die unterschiedlichen Erfolgsquoten ist in erster Linie darin zu sehen, dass das Suchverhalten der Prüfer mit eigenständiger Informationssuche durch eine geringere Breite gekennzeichnet ist.<sup>3</sup> Dieses deutet erneut darauf hin, dass eine adäquate Hypothesengenerierung aufgrund von potenziellen Defiziten im Suchverhalten nicht zwangsläufig bedeutet, dass der tatsächliche Abweichungsgrund letztendlich auch ausgewählt wird.

Darüber hinaus konnten die Autoren erkennen, dass Prüfer mit dem richtigen Hypothesensatz und ausgewogenen Testergebnissen signifikant erfolgreicher waren als Prüfer mit einem inkorrekten Hypothesensatz (d.h. der extern zur Verfügung gestellte Hypothesensatz umfasst nicht die tatsächliche Abweichungsursache) und ausgewogenen Testergebnissen.<sup>4, 5</sup> Dieses ist im Wesentlichen darauf zurückzuführen, dass letztgenannte Prüfergruppe nicht in der Lage war, eine neue Hypothese zu generieren, obwohl bereits Hinweise vorlagen, die auf die Gültigkeit einer Hypothese hindeuteten, die nicht im aktuellen Hypothesensatz vertreten war.<sup>6</sup> Dieses lässt vermuten, dass eine ausgewo-

---

<sup>1</sup> Vgl. im Folgenden ASARE/WRIGHT (2003), S. 242-244.

<sup>2</sup> Vgl. ASARE/WRIGHT (2003), S. 245 i.V.m. 247.

<sup>3</sup> Vgl. ASARE/WRIGHT (2003), S. 235 i.V.m. 246.

<sup>4</sup> Vgl. ASARE/WRIGHT (2003), S. 245 i.V.m. 247.

<sup>5</sup> Hierbei ist zu berücksichtigen, dass die Prüfer mit eigenständiger Hypothesengenerierung zu den Prüfern mit einem inkorrekten Hypothesensatz zugeordnet wurden, da lediglich einer von diesen im Stande war, einen Hypothesensatz zu generieren, der den tatsächlichen Grund beinhaltete. Vgl. ASARE/WRIGHT (2003), S. 246.

<sup>6</sup> Vgl. ASARE/WRIGHT (2003), S. 248.

gene Suchstrategie den Prüfer nicht grundsätzlich dazu befähigt, die Defizite einer anfänglichen Hypothesengenerierung zu kompensieren.<sup>1</sup> Demnach ist festzuhalten, dass für die Prüfungspraxis nicht nur eine breite bzw. ausgewogene Suchstrategie eine entscheidende Rolle für die erfolgreiche Durchführung von analytischen Prüfungshandlungen einnimmt, sondern auch eine durch iterative Hypothesengenerierung gekennzeichnete prüferische Vorgehensweise von besonderer Bedeutung ist.

#### **3.5.4.4 Einflussfaktoren der Informationssuche**

##### **3.5.4.4.1 Vorbemerkungen**

Folgt man generell der Auffassung, dass eine positive Suchstrategie aufgrund des möglichen Auftretens von konfirmatorischen Prozessen einerseits und des Übersehens gegenteiliger Informationen andererseits möglichst vermieden und vielmehr eine ausgewogene bzw. breite Informationssuche präferiert werden sollte, stellt sich in einem nächsten Schritt die Frage, welche Einflussfaktoren die Verwendung dieser Strategie fördern bzw. entgegenstehen.

Als mögliche externe Faktoren kommen dabei insbesondere Rechtfertigungsdruck, zeitliche Restriktionen sowie der Erhalt einer Mandantenerklärung in Betracht. Da es sich hierbei um prüfungsimmanente Faktoren handelt, die regelmäßig bei einer Jahresabschlussprüfung vorliegen,<sup>2</sup> kann die Untersuchung ihres Einflusses auf die Informationssuche einen Eindruck über die Vorgehensweise der Prüfungspraxis vermitteln. Ein potenzieller interner bzw. personenbezogener Einflussfaktor ist in der Erfahrung des Prüfers zu sehen.

##### **3.5.4.4.2 Rechtfertigungsdruck und zeitliche Beschränkungen**

ASARE/TROMPETER/WRIGHT untersuchten in ihre Studie den Einfluss von Rechtfertigungsdruck und zeitlichen Restriktionen.<sup>1</sup> Der Rechtfertigungsdruck wurde durch die Ankündigung eines Reviews durch einen Vorgesetzten induziert, zeitliche Restriktionen durch die Vorgabe eines Zeitbudgets, welches die Prüfer im Rahmen ihrer Informati-

---

<sup>1</sup> Vgl. ASARE/WRIGHT (2003), S. 235.

<sup>2</sup> Dass der Erhalt einer Mandantenerklärung gängige Praxis darstellt und zudem von den Prüfungsnormen verlangt wird, wurde bereits mehrfach erwähnt. Zur Prüfungsimmanenz von Rechtfertigungs- und Zeitdruck siehe ausführlich SCHREIBER (2000), S. 92-98.

onssuche nicht überschreiten durften.<sup>2</sup> Da beide Elemente regelmäßig im Rahmen einer Abschlussprüfung anzutreffen sind, kann vermutet werden, dass hierdurch die Prüfungspraxis widergespiegelt wird.

Die Ergebnisse wiesen darauf hin, dass sowohl Rechtfertigungsdruck als auch zeitliche Restriktionen einen signifikanten Einfluss auf den Umfang der durchgeführten Prüfungstests hatten. Prüfer mit Rechtfertigungsdruck führten durchschnittlich 6,15 Tests durch, Prüfer ohne Druck hingegen 5,08. Demgegenüber führte die Vorgabe eines Zeitbudgets dazu, dass Prüfer im Schnitt 4,83 Tests durchführten, Prüfer ohne zeitliche Restriktionen hingegen 6,51.<sup>3</sup>

Um festzustellen, welches konkrete prüferische Vorgehen für die Erhöhung bzw. Reduzierung des Umfangs verantwortlich war, untersuchten ASARE/TROMPETER/WRIGHT zudem, ob der Einfluss auf den Prüfungsumfang durch Änderungen in der Breite oder in der Tiefe der Suche determiniert wurde. Dabei stellten die Autoren fest, dass der durch Rechtfertigungsdruck erhöhte Prüfungsumfang im Wesentlichen durch eine Erhöhung der Anzahl der getesteten Hypothesen (Breite) verursacht wurde.<sup>4</sup> Prüfer mit Rechtfertigungsdruck testeten signifikant mehr Hypothesen als Prüfer ohne Druck (3,87 ggü. 3,02). Die Autoren führten dieses darauf zurück, dass verantwortliche Prüfer eher dazu geneigt waren, eine Vielzahl von potenziellen Hypothesen zu testen als Prüfer ohne Rechtfertigungsdruck.<sup>5</sup> Demgegenüber konnte kein signifikanter Einfluss von Rechtfertigungsdruck auf die Tiefe der Suche festgestellt werden (1,60 im Vgl. zu 1,74).<sup>6</sup>

Weiterhin stellten ASARE/TROMPETER/WRIGHT fest, dass der durch die Vorgabe einer zeitlichen Beschränkung reduzierte Prüfungsumfang im Wesentlichen durch eine geringere Tiefensuche verursacht wurde. Prüfer mit Zeitbudget führten im Schnitt signifikant weniger Tests pro Hypothese als Prüfer mit Zeitbudget durch (1,56 im Vgl. zu 1,79).<sup>7</sup> Keinen signifikanten Einfluss hatte die Vorgabe von Zeitbudgets hingegen auf die Breite der Suche (3,63 im Vgl. zu 3,29).<sup>8</sup> Dieser Befund steht im Einklang mit der sog. Filterstrategie, wonach potenzielle Hypothesen unter Zugrundelegung von weniger Nach-

---

<sup>1</sup> Eine ausführliche Darstellung des Versuchsaufbaus wurde bereits in Kap. 3.5.4.3 vorgenommen.

<sup>2</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 550-551.

<sup>3</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 552.

<sup>4</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 554.

<sup>5</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 552.

<sup>6</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 553.

<sup>7</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 554.

<sup>8</sup> Vgl. ASARE/TROMPETER/WRIGHT (2000), S. 552 i.V.m. 553.



weisen validiert bzw. falsifiziert werden.<sup>1</sup> „When there is no time budget, an auditor may, for example, desire two pieces of evidence to disconfirm a hypothesis. The first piece of evidence may be used to form beliefs and the second to increase confidence in the belief. Auditors under a time budget may be willing to make decisions at lower levels of confidence based on fewer pieces of evidence per hypothesis than when facing no time budget.”<sup>2</sup>

Zusammenfassend deuten die Ergebnisse von ASARE/TROMPETER/WRIGHT darauf hin, dass Rechtfertigungsdruck zu einer Verstärkung der Breiten- bzw. der ausgewogenen Suchstrategie führt und keinen wesentlichen Einfluss auf die Tiefe der Suche mit sich bringt. Zeitliche Beschränkungen hingegen führen zu einer Abnahme der Tiefensuche, ohne dabei eine Reduzierung der Breitensuche nach sich zu ziehen.

#### **3.5.4.4.3 Der Einfluss des Erhalts einer Mandantenerklärung auf die Eigenschaften der Informationssuche mit und ohne Berücksichtigung der branchenspezifischen Erfahrung**

##### **3.5.4.4.3.1 Ohne Berücksichtigung der Branchenerfahrung**

GREEN untersuchte in ihrer Studie, inwiefern der Erhalt einer Mandantenerklärung zu einer Beeinflussung der Informationssuche führt.<sup>3</sup> Dabei teilt sie die teilnehmenden Prüfer nach dem Erhalt einer Mandantenerklärung und nach dem Zeitpunkt des Erhalts in drei Gruppen auf. Anschließend konnten die Prüfer sequentiell Prüfungstests aus einer Liste mit 18 möglichen Tests auswählen und die Ergebnisse zur Kenntnis nehmen.<sup>4</sup>

GREEN konnte nachweisen, dass der Erhalt einer Mandantenerklärung keine signifikanten Auswirkungen auf den Umfang, die Breite und die Tiefe der Informationssuche ausübte.<sup>5</sup> Im Schnitt führten die beiden Prüfergruppen mit Mandantenerklärung 5,62 bzw. 5,90 Tests durch im Vergleich zu 6,15 Tests bei Prüfern ohne Mandantenerklärung. Weiterhin untersuchten Prüfer mit Mandantenerklärung durchschnittlich 3,86 bzw. 3,95 Hypothesen gegenüber 4,15 Hypothesen bei Prüfern ohne Mandantenerklärung. Auch hinsichtlich der Tiefe der Suche ergaben sich zwischen Prüfern mit bzw. ohne Mandan-

---

<sup>1</sup> Vgl. MCDANIEL (1990), S. 280-283 oder GLOVER (1997), S. 222-223.

<sup>2</sup> ASARE/TROMPETER/WRIGHT (2000), S. 552 i.V.m. 556

<sup>3</sup> Eine ausführliche Darstellung findet sich unter Kap. 3.5.3.3.2.1.2.1.

<sup>4</sup> Vgl. GREEN (2004a), S. 375-377. Dabei wurde allen Prüfern das gleiche Zeitbudget von 20 Stunden eingeräumt.

<sup>5</sup> Vgl. im Folgenden GREEN (2004a), S. 383.

tenerklärung keine signifikanten Unterschiede (1,36 bzw. 1,35 im Vgl. zu 1,40). Darüber hinaus ließen die Ergebnisse von GREEN keine erkennbaren Zusammenhänge zwischen dem Erhalt einer Mandantenerklärung und der Häufigkeit der durchgeführten Tests, die auf die inkorrekte Mandantenerklärung abzielten, erkennen.<sup>1</sup>

Hinsichtlich der Breite der Informationssuche lieferte eine Studie von ASARE/WRIGHT/WRIGHT vergleichbare Ergebnisse. Die Autoren ließen insgesamt 73 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 4,95 Jahren eine wesentliche Abweichung in der Bruttogewinnmarge untersuchen. Dabei wurde einer Prüfergruppe eine inkorrekte Mandantenerklärung zur Verfügung gestellt, den anderen Prüfern hingegen keine Mandantenerklärung bereitgestellt.<sup>2, 3</sup> Analog zu GREEN stellten die Autoren insgesamt keinen signifikanten Einfluss des Erhalts einer Mandantenerklärung auf die Breite der Informationssuche fest. Prüfer mit bzw. ohne Mandantenerklärung untersuchten durchschnittlich 3,1 bzw. 3,3 Hypothesen.<sup>4</sup> Allerdings erkannten ASARE/WRIGHT/WRIGHT im Gegensatz zu GREEN sehr wohl, dass Prüfer mit inkorrekt erhaltener Mandantenerklärung signifikant häufiger Tests durchführten, die auf die Mandantenerklärung abzielten, als Prüfer ohne Mandantenerklärung (1,1 im Vgl. zu 0,79 Test pro „vererbter“ Hypothese). Darüber hinaus stellten ASARE/WRIGHT/WRIGHT fest, dass Prüfer, denen eine inkorrekte Nicht-Fehlerhypothese mitgeteilt wurde, signifikant weniger Tests durchführten, die auf einen Fehler abzielten, als Prüfer ohne Mandantenerklärung (0,52 ggü. 0,74 pro Nicht-Fehlerursache).<sup>5</sup>

#### **3.5.4.4.3.2 Mit Berücksichtigung der Branchenerfahrung**

Eine Studie, die neben dem Erhalt einer inkorrekten Mandantenerklärung zudem den Einfluss der Branchenerfahrung auf die Informationssuche untersuchte, führte GREEN durch.<sup>6</sup> Dabei stellte GREEN fest, dass Branchenspezialisten durchschnittlich weniger Tests durchführten als Nicht-Spezialisten (5,55 ggü. 6,34).<sup>7</sup> Dieses konnte darauf zurückgeführt werden, dass Spezialisten eine geringe Breitensuche aufwiesen. Sie fokus-

---

<sup>1</sup> Vgl. GREEN (2004a), S. 383.

<sup>2</sup> Vgl. ASARE/WRIGHT/WRIGHT (1998), S. 19-20.

<sup>3</sup> Da es sich ansonsten um den identischen Versuchsaufbau wie bei ASARE/WRIGHT (2003) handelt, soll an dieser Stelle auf keine weiteren Details eingegangen werden. Siehe hierzu ausführlich die Versuchsbeschreibung unter Kap. 3.5.4.3.

<sup>4</sup> Vgl. ASARE/WRIGHT/WRIGHT (1998), S. 22-24.

<sup>5</sup> Vgl. ASARE/WRIGHT/WRIGHT (1998), S. 22-23.

<sup>6</sup> Für Details der Untersuchung soll auf Kap. 3.5.3.3.2.1.4.2 verwiesen werden.

<sup>7</sup> Vgl. GREEN (2004b), S. 13.

sierten im Schnitt 3,62 Hypothesen im Vergleich zu 4,31 Hypothesen bei Nicht-Spezialisten.<sup>1</sup> Zwar wurde zuvor festgehalten, dass Prüfer nach Möglichkeit eine Breiten- bzw. ausgewogene Suchstrategie anwenden sollten. Allerdings wiesen die Ergebnisse von GREEN darauf hin, dass die Branchenspezialisten signifikant erfolgreicher waren als Nicht-Spezialisten. Demnach scheint es bei der Anzahl der getesteten Hypothesen einen Trade-off zu geben, der eine optimale Suchbreite vermuten lässt.

Keinen signifikanten Einfluss hatte die Branchenerfahrung auf die Anzahl der Tests pro Hypothese (sog. Suchtiefe; 1,63 im Vgl. zu 1,52).<sup>2</sup> Dennoch konnte festgestellt werden, dass Spezialisten einen Test, der auf den tatsächlichen Abweichungsgrund gerichtet war, marginal häufiger wählten als Nicht-Spezialisten und Tests, die auf allgemeine Fehlerursachen bezogen waren, marginal seltener wählten.<sup>3</sup>

---

<sup>1</sup> Vgl. GREEN (2004b), S. 13.

<sup>2</sup> Vgl. GREEN (2004b), S. 13 i.V.m. 25.

<sup>3</sup> Vgl. GREEN (2004b), S. 25.

### 3.5.4.5 Tabellarische Darstellung der empirischen Befunde zur Phase der Informationssuche

| Autor(en)                                  | Teilnehmer |                                                     | Daten-<br>gewinnung | Daten-<br>auswertung                         | Ergebnisse                                                                                                                                                                                                                                                                                                            | Signifikanz         |
|--------------------------------------------|------------|-----------------------------------------------------|---------------------|----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
|                                            | Anzahl     | Art                                                 |                     |                                              |                                                                                                                                                                                                                                                                                                                       |                     |
| KAPLAN/RECKERS<br>(1989)                   | 71         | k.A.<br>Ø = 8 J.                                    | Experiment          | Regressionsanalyse<br><br>Regressionsanalyse | Unerfahrenere Prüfer (Ø < 5 J.) führen wesentlich mehr Tests durch, die auf die Überprüfung ihrer generierten Hypothese gerichtet ist (sog. positive Suchstrategie bzw. confirmatorische Prozesse).<br>Erfahrenere Prüfer (Ø > 6 J.) weisen hingegen keine positive Suchstrategie bzw. confirmatorische Prozesse auf. | 0,0013<br><br>0,847 |
| BEDARD/BIGGS<br>(1991a)                    | 21         | 10 Manager<br>k.A.<br><br>11 Prüfungsleiter<br>k.A. | Experiment          | Verbale Protokolle                           | 38,1% der Prüfer konnten ihre selbstgenerierte Hypothese nicht verwerfen, obwohl bereits Hinweise vorlagen, die die Gültigkeit ihrer Hypothese widerlegten.                                                                                                                                                           | k.A.                |
| HEIMAN-HOFFMAN/<br>MOSEER/JOSEPH<br>(1995) | 54         | k.A.<br>Ø = 3 J.                                    | Experiment          | k.A.                                         | Prüfer haben Schwierigkeiten, ihre selbstgenerierte Hypothese zu verwerfen, auch wenn Hinweise vorliegen, die die Gültigkeit ihrer Hypothese widerlegen.                                                                                                                                                              | k.A.                |
| ANDERSON/KOONCE<br>(1995)                  | 60         | Prüfungsleiter<br>Ø = 2 J.                          | Experiment          | Wilcoxon-Vorzeichen-<br>Rang-Test            | Prüfer, die eine inkorrekte Mandantenerklärung erhielten, konnten signifikant häufiger Argumente für die Gültigkeit der Mandantenerklärung liefern, als Prüfer mit korrekter Mandantenerklärung keine Argumente für die Gültigkeit der Mandantenerklärung liefern konnten.                                            | 0,0001              |
| BIERSTAK-<br>KER/BEDARD/BIGGS<br>(1999)    | 12         | Prüfungsleiter<br>3 bis 5 J.                        | Experiment          | Verbale Protokolle                           | Der Erhalt einer Mandantenerklärung führt dazu, dass Prüfer Hinweise fokussieren, die die Mandantenerklärung bestätigen.                                                                                                                                                                                              | k.A.                |
| PETERSON/WONG-<br>ON-WING (2000)           | 93         | Prüfungsleiter<br>Ø = 3,3 J.                        | Experiment          | t-Test                                       | Prüfer, die die eingebettete Version bearbeiteten und eine positive Suchstrategie anwendeten, wa-                                                                                                                                                                                                                     | 0,067               |

|                              |    |                              |            |                                        |                                                                                                                                                                                                                                                                                                                                                                                          |                                              |
|------------------------------|----|------------------------------|------------|----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
|                              |    |                              |            | t-Test                                 | ren marginal weniger effizient, als Prüfer, die ebenfalls die eingebettete Version bearbeiteten und keine positive Suchstrategie anwendeten. Prüfer, die die umschlossene Version bearbeiteten und eine positive Suchstrategie anwendeten, waren signifikant erfolgreicher, als Prüfer, die ebenfalls die umschlossene Version bearbeiteten und keine positive Suchstrategie anwendeten. | 0,009                                        |
| CHURCH (1991)                | 79 | k.A.<br>Ø = 2,5 J.           | Experiment | Varianzanalyse/<br>Mann-Whitney-U-Test | Prüfer mit einer starken Bindung (commitment) zur Hypothese fokussieren den entsprechenden Transaktionskreis signifikant stärker als Prüfer ohne Bindung.                                                                                                                                                                                                                                | 0,0396 <sup>1</sup> /<br>0,0066 <sup>2</sup> |
|                              |    |                              |            | Varianzanalyse/<br>Mann-Whitney-U-Test | Prüfer mit einer starken Bindung (commitment) zur Hypothese messen Informationen, die auf Probleme im entsprechenden Transaktionskreis deuten (sog. confirming evidence) eine signifikant stärkere Bedeutung zu als Prüfer ohne Bindung.                                                                                                                                                 | 0,0347/0,0307                                |
|                              |    |                              |            | Varianzanalyse/<br>Mann-Whitney-U-Test | Prüfer mit einer starken Bindung (commitment) zur Hypothese messen Informationen, die auf Probleme in einem anderen Transaktionskreis deuten (sog. disconfirming evidence) keine signifikant geringere Bedeutung zu als Prüfer ohne Bindung.                                                                                                                                             | 0,178/0,2119                                 |
| PETERSON/WONG-ON-WING (2000) | 93 | Prüfungsleiter<br>Ø = 3,3 J. | Experiment | Chi-Quadrat-Test                       | Die Zuteilung der Suchstrategie nach der aktuellen Hypothese führt im Vgl. zur Zuteilung nach der anfänglichen Hypothese zu einer signifikant                                                                                                                                                                                                                                            | 0,0325                                       |

<sup>1</sup> Signifikanzniveau der Varianzanalyse.

<sup>2</sup> Signifikanzniveau des Mann-Whitney-U-Tests.

|                               |     |                     |            |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                 |
|-------------------------------|-----|---------------------|------------|------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
|                               |     |                     |            |                                                                        | höheren Anzahl von durchgeführten Tests, die dem ausgewählten Transaktionskreis zuzuordnen sind (sog. positive Suchstrategie) und zu einer signifikant niedrigeren Anzahl von Tests, die einem anderen Transaktionskreis anzurechnen sind (sog. negative Suchstrategie).                                                                                                                                                             |                                 |
| ASARE/TROMPETER/WRIGHT (2000) | 91  | k.A.<br>Ø = 2,9 J.  | Experiment | Chi-Quadrat-Test<br><br>Chi-Quadrat-Test                               | Die Breite der Informationssuche hat einen signifikant positiven Einfluss auf die Erfolgsquote.<br>Bei der Tiefe der Informationssuche konnte hingegen kein signifikanter Einfluss auf die Erfolgsquote festgestellt werden.                                                                                                                                                                                                         | 0,0005<br><br>0,5022            |
| ASARE/WRIGHT (2003)           | 102 | k.A.<br>Ø = 2,05 J. | Experiment | t-Test                                                                 | Erfolgreiche Prüfer testen signifikant mehr Hypothesen (Breite der Suche) als nicht erfolgreiche Prüfer.                                                                                                                                                                                                                                                                                                                             | 0,0154                          |
| GREEN/TROTMAN (2003)          | 82  | k.A.<br>Ø = 3,8 J.  | Experiment | Varianzanalyse<br><br>t-Test<br><br>Varianzanalyse                     | Die Tiefensuche hat einen signifikanten Einfluss auf den Erfolg der Prüfer.<br>Nicht erfolgreiche Prüfer weisen eine signifikant stärker Tiefensuche auf als erfolgreiche Prüfer.<br>Erfolgreiche Prüfer testen nicht signifikant mehr Hypothesen als nicht erfolgreiche Prüfer.                                                                                                                                                     | 0,046<br><br>0,000<br><br>0,539 |
| ASARE/WRIGHT (2003)           | 102 | k.A.<br>Ø = 2,05 J. | Experiment | Wilcoxon-Vorzeichen-Rang-Test<br><br><br>Wilcoxon-Vorzeichen-Rang-Test | Prüfer mit dem richtigen Hypothesensatz und ausgewogenen Testergebnissen waren marginal erfolgreicher als Prüfer, die ebenfalls den richtigen Hypothesensatz erhielten, aber eine eigenständige Hypothesensuche durchführten.<br>Prüfer mit dem richtigen Hypothesensatz und ausgewogenen Testergebnissen waren signifikant erfolgreicher als Prüfer, die einen inkorrekten Hypothesensatz und ausgewogene Testergebnisse erhielten. | 0,0643<br><br><br>0,00212       |

|                                    |    |                               |            |                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                |
|------------------------------------|----|-------------------------------|------------|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|
| ASARE/TROMPE-<br>TER/WRIGHT (2000) | 91 | k.A.<br>$\emptyset = 2,9$ J.  | Experiment | Varianzanalyse                    | Rechtfertigungsdruck und zeitliche Restriktionen haben einen signifikanten Einfluss auf den Umfang der durchgeführten Prüfungstests.<br>Rechtfertigungsdruck hat einen signifikanten Einfluss auf die Breite der Informationssuche und keinen signifikanten Einfluss auf die Tiefe der Suche.<br>Zeitliche Restriktionen haben keinen signifikanten Einfluss auf die Breite der Suche, aber einen signifikanten Einfluss auf die Tiefensuche. | 0,008 <sup>1</sup> /0,001 <sup>2</sup>                         |
|                                    |    |                               |            | Varianzanalyse                    |                                                                                                                                                                                                                                                                                                                                                                                                                                               | 0,0002/0,1314                                                  |
|                                    |    |                               |            | Varianzanalyse                    |                                                                                                                                                                                                                                                                                                                                                                                                                                               | 0,9902/0,0145                                                  |
| GREEN (2004a)                      | 63 | k.A.<br>$\emptyset = 3,84$ J. | Experiment | Varianzanalyse                    | Der Erhalt einer Mandantenerklärung hat keinen signifikanten Einfluss auf den Umfang, die Breite und die Tiefe der Suche.<br>Die Ergebnisse lassen keinen eindeutigen Zusammenhang zwischen dem Erhalt einer Mandantenerklärung und der Anzahl der durchgeführten Tests, die auf die Überprüfung der Mandantenerklärung gerichtet sind, erkennen.                                                                                             | 0,335 <sup>3</sup> bzw.<br>0,267 <sup>4</sup> /0,521/<br>0,987 |
|                                    |    |                               |            | Chi-Quadrat-Test                  |                                                                                                                                                                                                                                                                                                                                                                                                                                               | 0,027 <sup>5</sup>                                             |
| ASA-<br>RE/WRIGHT/WRIGHT<br>(1998) | 73 | k.A.<br>$\emptyset = 4,95$ J. | Experiment | t-Test                            | Der Erhalt einer Mandantenerklärung hat keinen signifikanten Einfluss auf die Breite der Suche.<br>Der Erhalt einer Mandantenerklärung hat einen signifikanten Einfluss auf die Anzahl der durchgeführten Tests, die auf die Überprüfung der Mandantenerklärung gerichtet sind.                                                                                                                                                               | 0,1878                                                         |
|                                    |    |                               |            | Wilcoxon-Vorzeichen-<br>Rang-Test |                                                                                                                                                                                                                                                                                                                                                                                                                                               | 0,0174                                                         |

<sup>1</sup> Signifikanzniveau für Rechtfertigungsdruck.

<sup>2</sup> Signifikanzniveau für zeitliche Restriktionen.

<sup>3</sup> Signifikanzniveau bei Anzahl der durchgeführten Tests als Maß für den Umfang der Suche.

<sup>4</sup> Signifikanzniveau bei in Anspruch genommener Prüfungszeit als Maß für den Umfang der Suche.

<sup>5</sup> Prüfer, die die Mandantenerklärung nach einer eigenständigen Hypothesengenerierung erhielten, testeten diese signifikant seltener als Prüfer, die die Mandantenerklärung vor der eigenständigen Hypothesengenerierung erhielten, bzw. Prüfer ohne Mandantenerklärung.

|               |    |                  |            |                                  |                                                                                                                                                                                             |                                        |
|---------------|----|------------------|------------|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
|               |    |                  |            | Wilcoxon-Vorzeichen-Rang-Test    | Prüfer, die eine Nicht-Fehlererklärung vom Mandanten erhalten, führten signifikant weniger Tests durch, die auf einen Nicht-Fehler ausgerichtet sind, als Prüfer ohne Mandantenerklärung.   | 0,0322                                 |
| GREEN (2004b) | 58 | k.A.<br>Ø = 4 J. | Experiment | Varianzanalyse                   | Branchenspezialisten führen signifikant weniger Tests durch als Nicht-Spezialisten.                                                                                                         | 0,023                                  |
|               |    |                  |            | Varianzanalyse                   | Branchenspezialisten untersuchen signifikanter weniger Hypothesen als Nicht-Spezialisten.                                                                                                   | 0,023                                  |
|               |    |                  |            | Chi-Quadrat-Test                 | Branchenspezialisten sind signifikant erfolgreicher als Nicht-Spezialisten.                                                                                                                 | 0,019                                  |
|               |    |                  |            | Varianzanalyse                   | Branchenerfahrung hat keinen signifikanten Einfluss auf die Tiefensuche.                                                                                                                    | 0,311                                  |
|               |    |                  |            | Varianzanalyse, Chi-Quadrat-Test | Branchenspezialisten führen marginal weniger Tests durch, die auf allgemeine Fehlerursachen gerichtet sind, und marginal mehr Tests, die auf die korrekte Abweichungsursache gerichtet ist. | 0,058 <sup>1</sup> /0,060 <sup>2</sup> |

Tabelle 4: Ergebnisse der empirischen Prüfungsforschung zur Phase der Informationssuche

<sup>1</sup> Signifikanzniveau hinsichtlich der Durchführung allgemeiner Tests (Varianzanalyse).

<sup>2</sup> Signifikanzniveau hinsichtlich der Durchführung eines Tests, der auf die korrekte Abweichungsursache gerichtet ist (Chi-Quadrat-Test).



### **3.5.5 Studien zur Phase der Hypothesenbewertung**

#### **3.5.5.1 Vorbemerkung**

Nachdem der Prüfer seine Hypothesen generiert und die (hypothesengesteuerte) Suche nach verifizierenden bzw. falsifizierenden Informationen abgeschlossen hat, obliegt es ihm, sich für eine Hypothese oder mehrere Hypothesen als (unterstellte) Ursache für die identifizierte Abweichung zu entscheiden. Dabei werden die generierten Hypothesen vor dem Hintergrund der erlangten Informationen einer Wahrscheinlichkeitsbeurteilung unterzogen und die Hypothese mit der höchsten Wahrscheinlichkeit selektiert.

Dabei interessierte sich die empirische Prüfungsforschung u.a. für die generelle Erwartungshaltung von Prüfern gegenüber festgestellten Abweichungen. Hier stellt sich die Frage, ob Prüfer tendenziell davon ausgehen, dass eine Abweichung durch einen Fehler oder durch einen Nicht-Fehler verursacht wurde. Im Mittelpunkt der Forschungsbemühungen stand zum einen jedoch die Untersuchung des Einflusses einer „vererbten“ Abweichungsursache (etwa durch den Mandanten oder durch eine EDV-gestützte Entscheidungshilfe) auf die Hypothesenbewertung und zum anderen die Frage nach der Bewertung von multiplen Abweichungshypothesen.

#### **3.5.5.2 Generelle Erwartungshaltung von Prüfern gegenüber festgestellten Abweichungen**

Wahrscheinlichkeitsbeurteilungen werden wesentlich von den generellen Erwartungshaltungen der jeweiligen Prüfer beeinflusst. In Bezug auf die Beurteilung einer durch analytische Prüfungshandlungen identifizierten Abweichung stellt sich diesbezüglich die Frage, ob Prüfer tendenziell dazu neigen, einen Fehler oder einen Nicht-Fehler als Abweichungsursache zu unterstellen. Durch eine derartige Erwartungshaltung könnte die Hypothesenbewertung dahingehend verzerrt werden, dass der Prüfer seine Wahrscheinlichkeitsbeurteilungen nicht auf der Basis der objektiv zu Verfügung stehenden Informationen fällt, sondern unabhängig von der Informationslage tendenziell zur Annahme einer Fehler- bzw. Nicht-Fehlerhypothese neigt.

Um diese Frage genauer zu durchleuchten, führten KAPLAN/RECKERS eine Studie durch, in der sie 71 teilnehmende Prüfer nach ihrer generellen Erwartungshaltung hinsichtlich

des Grunds einer unerwarteten wesentlichen Abweichung fragten.<sup>1</sup> Dabei stellten die Autoren fest, dass die Prüfererwartungen im Wesentlichen von der Prüfungserfahrung abhängen. Prüfer mit weniger Berufserfahrung neigten verstärkt zur Annahme, dass die Abweichung durch einen Fehler in der Rechnungslegung verursacht wurde. Erfahrene Prüfer hingegen tendieren eher dazu, die festgestellte Abweichung auf einen Nicht-Fehler zurückzuführen.<sup>2</sup>

KAPLAN/RECKERS führten diese Beobachtung weitestgehend darauf zurück, dass die Gegebenheiten in der Realität erheblich von den ausbildungsbedingten Vorstellungen über die Prüfungspraxis abweichen.<sup>3</sup> Die berufliche Ausbildung betont überwiegend, dass der Zweck der Jahresabschlussprüfung in der Aufdeckung von Fehlern in der Rechnungslegung des Mandanten zu sehen ist. Dieses könnte dazu führen, dass die Vorstellung von Berufsanfängern bzw. unerfahrenen Prüfern davon dominiert wird, dass unerwartete Diskrepanzen in den Jahresabschlussdaten überwiegend durch Fehler hervorgerufen werden. Prüfer mit zunehmender Berufspraxis machen jedoch die Erfahrung, dass eine Vielzahl von Jahresabschlüssen nur in einem geringen Maß von Jahresabschlussfehlern betroffen ist, so dass sie ihre Erwartungen hinsichtlich der Ursache einer festgestellten Abweichung sukzessive anpassen.<sup>4</sup>

### **3.5.5.3 Die Bewertung einer „vererbten“ Hypothese**

#### **3.5.5.3.1 Der Erhalt einer Abweichungshypothese durch den Mandanten bzw. dessen Mitarbeiter**

##### **3.5.5.3.1.1 Der Einfluss des Erhalts einer Mandantenerklärung auf die Hypothesenbewertung**

Wie bereits mehrfach betont, ist der Frage, ob die durch ISA 520.18 sowie IDW PS 312.27 nahe gelegte Einholung einer Mandantenerklärung eine potenzielle Gefahr für

---

<sup>1</sup> Eine detaillierte Beschreibung des Untersuchungsaufbaus findet sich in Kap. 3.5.4.2.1.

<sup>2</sup> Vgl. KAPLAN/RECKERS (1989), S. 545.

<sup>3</sup> Vgl. im Folgenden KAPLAN/RECKERS (1989), S. 549.

<sup>4</sup> Die Vermutung, dass Prüfer im Rahmen ihrer beruflichen Tätigkeiten nur selten mit fehlerbedingten Abweichungen konfrontiert werden, konnte ebenfalls durch die empirisch-archivistischen Untersuchungen von WRIGHT/ASHTON (1989) sowie ASHTON (1991) bestätigt werden. In der Folgezeit konnte die empirische Prüfungsforschung regelmäßig feststellen, dass die teilnehmenden Prüfer eine zu beurteilende Diskrepanz im Jahresabschluss überwiegend auf einen Nicht-Fehler zurückführten. GREEN/TROTMAN (2003), S. 227 etwa konnten nachweisen, dass knapp 71% ihrer Teilnehmer einen Nicht-Fehlergrund selektierten. Ein vergleichbares Ergebnis ermittelte GREEN (2004a), S. 386. Im

die Effektivität und Effizienz von Jahresabschlussprüfungen darstellt, eine besondere Bedeutung beizumessen.<sup>1</sup> Ein großer Teil der empirischen Prüfungsforschung widmete sich der Frage, inwiefern die Verfügbarkeit einer Mandantenerklärung die anschließende eigene Generierung von Fehlerhypothesen beeinflusst (sog. interference effect). Diesbezüglich konnten jedoch keine eindeutigen Nachweise erbracht werden, so dass in diesem Punkt generell kein negativer Einfluss auf die Jahresabschlussprüfung konstatiert werden kann.<sup>2</sup>

Der interference effect ist jedoch nur eine Möglichkeit, die prüferische Urteilsbildung durch die Verfügbarkeit einer Mandantenerklärung negativ zu beeinflussen. Eine weitere Möglichkeit ergibt sich in der Phase der Hypothesenbewertung. Berücksichtigt man, dass Prüfer tendenziell dazu neigen, festgestellte Abweichungen auf Nicht-Fehlerursachen zurückzuführen,<sup>3</sup> und Mandanten bei entsprechender Nachfrage durch den Jahresabschlussprüfer generell den Anreiz mitbringen, auf einen Nicht-Fehler zu verweisen,<sup>4</sup> erscheint eine potenzielle Gefahr für die Effektivität und Effizienz von Jahresabschlussprüfungen durchaus immanent.

In diesem Zusammenhang ist die Untersuchung von GREEN/TROTMAN zu betrachten.<sup>5</sup> Sie fanden heraus, dass 34% der teilnehmenden Prüfer der erhaltenen Mandantenerklärung die höchste Wahrscheinlichkeit zuordneten, obwohl die eingeholten Informationen die Inkorrektheit dieser Mandantenerklärung bereits deutlich erkennen ließen.<sup>6, 7</sup> Noch deutlicher fielen die Ergebnisse in der Untersuchung von GREEN aus.<sup>8</sup> Sie stellte fest, dass 42% der Prüfer die Mandantenerklärung akzeptierten, obwohl die erlangten Prüfungsnachweise auch hier auf die Inkorrektheit der Mandantenerklärung hindeuteten.<sup>9</sup> Um einen Vergleichsmaßstab zu erhalten, ließ GREEN im Gegensatz zu GREEN/TROTMAN weitere 20 Prüfer den identischen Fall allerdings ohne Erhalt einer inkorrekten Mandantenerklärung bearbeiten. Von diesen 20 Prüfern wählte nicht einer die Mandantenerklärung als wahrscheinlichsten Grund für die festgestellte Abweichung

---

Rahmen ihrer Untersuchung wählten 74% der Prüfer eine Nicht-Fehlerursache als tatsächlichen Grund für die identifizierte Abweichung aus.

<sup>1</sup> Siehe Kap. 3.5.2.3 und insbesondere Kap. 3.5.3.3.2.1.1.

<sup>2</sup> Vgl. hierzu ausführlich Kap. 3.5.3.3.2.1.2.1 sowie Kap. 3.5.3.3.2.1.2.3.

<sup>3</sup> Siehe hierzu ausführlich Kap. 3.5.5.2.

<sup>4</sup> Vgl. hierzu Kap. 3.5.3.3.2.1.1.

<sup>5</sup> Eine ausführliche Beschreibung der Untersuchung findet sich in Kap. 3.5.3.2.

<sup>6</sup> Vgl. GREEN/TROTMAN (2003), S. 227.

<sup>7</sup> Ihre Wahrscheinlichkeitseinschätzungen fielen dabei signifikant höher aus als die der anderen Prüfer. Vgl. GREEN/TROTMAN (2003), S. 233.

<sup>8</sup> Für eine detaillierte Darstellung der Studie soll auf Kap. 3.5.3.3.2.1.2.1 verwiesen werden.

<sup>9</sup> Vgl. GREEN (2004a), S. 386.

aus.<sup>1, 2</sup> Wenn auch nicht derart eindeutig wie bei GREEN, weisen die Ergebnisse von ASARE/WRIGHT/WRIGHT in die gleiche Richtung.<sup>3</sup> Sie erkannten, dass 67,5% der Prüfer, die eine inkorrekte Mandantenerklärung erhielten, diese auch als wahrscheinlichste Ursache für die festgestellte Abweichung selektierten. Prüfer ohne Mandantenerklärung wählten diesen Nicht-Fehlergrund hingegen nur in 43,7% der Fälle aus.<sup>4, 5</sup> Insgesamt können diese Ergebnisse als eindeutiger Nachweis für den verzerrenden Einfluss des Erhalts einer Mandantenerklärung gesehen werden.

### **3.5.5.3.1.2 Verzerrungen im Rahmen der Hypothesenbewertung sowie potenzielle debiasing-Mechanismen**

#### **3.5.5.3.1.2.1 Der Einfluss der Dokumentation der Mandantenerklärung**

##### **3.5.5.3.1.2.1.1 Vorbemerkungen**

Individuen neigen dazu, einem möglichen Ereignis eine höhere Eintrittswahrscheinlichkeit zu zurechnen, wenn dieses potenzielle Ereignis schriftlich festgehalten wurde. Dieses in der psychologischen Forschung nachgewiesene Phänomen wird auch als sog. *written explanation effect* bezeichnet.<sup>6</sup> Darüber hinaus konstatiert die psychologische Forschung, dass der *written explanation effect* Individuen dazu veranlasst, an ihrer Wahrscheinlichkeitseinschätzung festzuhalten, auch wenn gegenteilige Informationen erlangt werden.<sup>1</sup>

Aufgrund der im Rahmen der Abschlussprüfung vorherrschenden Dokumentationspflichten ist nicht auszuschließen, dass Abschlussprüfer ebenfalls einer derartigen Verzerrung ausgesetzt sind. Kritisch wäre ein derartiger Einfluss auf die prüferische Urteilsbildung dann, wenn zwar eine plausible Erklärung dokumentiert wird, diese Erklärung jedoch nicht oder nur zum Teil für den untersuchten Aspekt verantwortlich ist. Ein derartiger Sachverhalt kommt insbesondere für die Durchführung von analytischen Prüfungshandlungen in Frage. Wie bereits mehrfach erwähnt, tendieren Prüfer dazu, den Mandanten als erste Informationsquelle für potenzielle Gründe zu konsultieren. Darüber

---

<sup>1</sup> Vgl. GREEN (2004a), S. 386.

<sup>2</sup> Demnach wählten Prüfer ohne Mandantenerklärung diese signifikant seltener aus als Prüfer mit Mandantenerklärung. Vgl. GREEN (2004a), S. 387.

<sup>3</sup> Für eine ausführliche Darstellung des Untersuchungsaufbaus sei auf Kap. 3.5.4.4.3.1 verwiesen.

<sup>4</sup> Vgl. ASARE/WRIGHT/WRIGHT (1998), S. 22.

<sup>5</sup> Demnach wählten Prüfer mit Mandantenerklärung diese signifikant häufiger als Fehlerursache aus als Prüfer ohne Mandantenerklärung. Vgl. ASARE/WRIGHT/WRIGHT (1998), S. 22.

<sup>6</sup> Vgl. ANDERSON/SECHLER (1986), S. 24-34.

hinaus wird eine derartige Vorgehensweise auch von den geltenden Prüfungsstandards nahe gelegt.<sup>2</sup> Mandanten hingegen bringen den Anreiz mit, zunächst eine plausible jedoch nicht richtige oder unvollständige Nicht-Fehlerbegründung zu liefern.<sup>3</sup> Sollte der Prüfer diese in Erfüllung seiner Dokumentationspflichten schriftlich niederlegen, so bestünde die Gefahr, dass er dieser eine höhere Wahrscheinlichkeit zuordnet und anschließend als richtige Begründung akzeptiert, selbst dann, wenn während des fortlaufenden Prüfungsprozesses gegenteilige Informationen zu Tage treten.

Studien, die diese potenzielle Verzerrung für den Prüfungsbereich untersuchten, führten ANDERSON/KOONCE und KOONCE durch. Die Studien unterschieden sich allerdings dahingehend voneinander, dass Erstere innerhalb ihres Untersuchungsaufbaus keine vollständig inkorrekte Mandantenerklärung zur Verfügung stellte. Vielmehr erhielten die Prüfer eine Mandantenerklärung, die teilweise korrekt war, d.h. die Mandantenerklärung konnte die festgestellte Abweichung nicht vollständig, sondern nur teilweise erklären. Die Autoren beriefen sich auf COGLITORE/BERRYMAN, die im Rahmen einer empirisch-archivistischen Studie aufzeigten, dass unerwartete wesentliche Abweichungen in den Jahresabschlussdaten häufig durch mehrere Ursachen hervorgerufen werden.<sup>4</sup> Zudem stellten HIRST/KOONCE in einer Felduntersuchung fest, dass Prüfer regelmäßig von lediglich einem Abweichungsgrund ausgehen und ihre Untersuchung häufig zu früh abbrechen, ohne die Abweichungsursachen vollständig ergründet zu haben.<sup>5</sup> Für die Prüfungspraxis könnte dieses ebenso wie die Annahme einer inkorrekten Mandantenerklärung eine ernsthafte Beeinträchtigung der Prüfungsqualität nach sich ziehen, da eine korrekte aber unvollständige Mandantenerklärung frühzeitig akzeptiert werden könnte und Fehler oder Unregelmäßigkeiten im Jahresabschluss unentdeckt blieben.<sup>1</sup>

#### **3.5.5.3.1.2.1.2 Einfluss einer teilweise korrekten Mandantenerklärung**

Eine der ersten Studien, die den Einfluss einer teilweise korrekten Mandantenerklärung auf die Hypothesenbewertung untersuchte, führten ANDERSON/KOONCE durch. Sie ließen 60 Prüfer mit einer durchschnittlichen Prüfungserfahrung von zwei Jahren eine unerwartete wesentliche Abweichung in der Bruttogewinnmarge analysieren. Den teil-

---

<sup>1</sup> Vgl. ANDERSON/LEPPER/ROSS (1980), S. 1034-1049.

<sup>2</sup> Vgl. IDW PS 312.27, ISA 520.18.

<sup>3</sup> So auch KOONCE (1992), S. 63.

<sup>4</sup> Vgl. COGLITORE/BERRYMAN (1988), S. 160-161.

<sup>5</sup> Vgl. HIRST/KOONCE (1996), S. 471, 475.

nehmenden Prüfern wurde mitgeteilt, dass der Mandant diese Abweichung mit einer Änderung im Verkaufssortiment begründet. Die Autoren teilten die Prüfer in zwei Gruppen ein, wobei den Fallunterlagen der einen Prüfergruppe entnommen werden konnte, dass die Mandantenerklärung vollständig für die festgestellte Abweichung verantwortlich war. Der anderen Prüfergruppe wurden hingegen Informationen zur Verfügung gestellt, die erkennen ließen, dass die Mandantenerklärung lediglich 40% der Abweichung erklärte. Daraufhin wurden alle Prüfer aufgefordert, eine schriftliche Erklärung über die Gültigkeit der Mandantenerklärung abzugeben und anschließend eine Wahrscheinlichkeit dafür zu treffen, dass mind. 90% der Abweichung durch die Mandantenerklärung begründet werden kann.<sup>2</sup>

ANDERSON/KOONCE stellten fest, dass 20 der 30 Prüfer mit der teilweise korrekten Mandantenerklärung dazu in der Lage waren, Argumente für die Richtigkeit und Vollständigkeit der Mandantenerklärung zu dokumentieren. Diese schätzten die Wahrscheinlichkeit der Richtigkeit und Vollständigkeit der Mandantenerklärung auf 63,31% und damit vergleichbar so hoch wie 28 der 30 Prüfer mit der vollständig korrekten Mandantenerklärung (63,57%), die ebenfalls dazu in der Lage waren, Argumente für die Richtigkeit und Vollständigkeit der Mandantenerklärung schriftlich festzuhalten.<sup>3</sup> Die Autoren führten die hohe Wahrscheinlichkeitseinschätzung der Prüfer mit teilweise korrekter Mandantenerklärung auf den *written explanation effect* zurück.

Weitere Analysen ließen jedoch erkennen, dass 19 der 20 Prüfer, die die Wahrscheinlichkeit der Richtigkeit und Vollständigkeit der Mandantenerklärung hoch einschätzten, obwohl sie nur für einen Teil der Abweichung ursächlich war, lediglich die Plausibilität der Mandantenerklärung überprüften. Nicht vorgenommen wurde hingegen eine Quantifizierung der Auswirkungen, die Gewissheit darüber verschaffen kann, ob die Abweichung auch vollständig durch die Mandantenerklärung verursacht wurde.<sup>4</sup> Darüber hinaus konnte festgestellt werden, dass acht der zehn Prüfer, die die Unvollständigkeit der Mandantenerklärung erkannten, eine rechnerische Überprüfung vornahmen, so dass die

---

<sup>1</sup> Vgl. ANDERSON/KOONCE (1998), S. 4. Ähnlich argumentieren ANDERSON/KOONCE (1995), S. 125 sowie ANDERSON/KOONCE (1998), S. 2.

<sup>2</sup> Vgl. ANDERSON/KOONCE (1995), S. 126-127.

<sup>3</sup> Weshalb Prüfer weiterhin an einer Hypothese festhalten, obwohl die Informationslage bereits gegen deren Gültigkeit spricht, wurde bereits unter Kap. 3.5.4.2.2 ausführlich untersucht.

<sup>4</sup> Vgl. ANDERSON/KOONCE (1995), S. 129-130.

Autoren vermuteten, dass die Quantifizierung einer Mandantenerklärung ein probates Mittel darstellt, um dem *written explanation effect* entgegenzuwirken.<sup>1</sup>

### 3.5.5.3.1.2.1.3 Einfluss einer vollständig inkorrekten Mandantenerklärung

Um zu untersuchen, ob die Hypothesenbewertung von Prüfern auch bei Erhalt einer vollständig inkorrekten Mandantenerklärung durch den *written explanation effect* gekennzeichnet ist, ließ KOONCE 26 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 2,8 Jahren einen vergleichbaren Fall wie ANDERSON/KOONCE bearbeiten. Den teilnehmenden Prüfern wurde mitgeteilt, dass der Mandant diese Abweichung mit einer Änderung im Verkaufssortiment begründet. Die Fallunterlagen beinhalteten zwar Informationen, welche die vom Mandanten angegebene Sortimentsänderung bestätigten, allerdings ging aus diesen Unterlagen nicht eindeutig hervor, dass die Mandantenerklärung tatsächlich zu der festgestellten Abweichung geführt hat.<sup>2</sup> Nach Erhalt der Mandantenerklärung wurde die Hälfte der Prüfer aufgefordert, Argumente zu dokumentieren, die für die Richtigkeit der Mandantenerklärung sprechen (treatment group). Die andere Hälfte der Prüfer hingegen nicht (control group). Anschließend sollten alle Prüfer eine Wahrscheinlichkeitsbeurteilung für die Richtigkeit der Mandantenerklärung abgeben.<sup>3</sup> Die Ergebnisse zeigten, dass die Wahrscheinlichkeitseinschätzung der treatment group mit 76,62% wesentlich größer war als die 64,62% der control group.<sup>4</sup> Dieser Einfluss der Dokumentation auf die Beurteilung von Hypothesen konnte in einer vergleichbaren Studie von KOONCE/PHILLIPS ebenfalls nachgewiesen werden.<sup>5</sup> Im Unterschied zu KOONCE wurden bei KOONCE/PHILLIPS nicht die Wahrscheinlichkeitsurteile von zwei unterschiedlichen Gruppen miteinander verglichen (sog. between-subject design), sondern die Beurteilung der identischen Prüfer zu unterschiedlichen Zeitpunkten (sog. within-subject design), d.h. vor und nach der schriftlichen Dokumentation.<sup>6</sup> Es stellte sich heraus, dass die Prüfer vor der Dokumentation die Wahrscheinlichkeit der Mandan-

---

<sup>1</sup> Vgl. ANDERSON/KOONCE (1995), S. 130.

<sup>2</sup> Vgl. KOONCE (1992), S. 65.

<sup>3</sup> Vgl. KOONCE (1992), S. 65-66.

<sup>4</sup> Vgl. KOONCE (1992), S. 65-66. Darüber hinaus wurde dieser signifikante Einfluss nicht nur zwischen den beiden unterschiedlichen Gruppen (between subject design) festgestellt, sondern auch bei den einzelnen Prüfern der control group (within-subject design). KOONCE forderte Letztere nach ihrer Wahrscheinlichkeitseinschätzung ebenfalls zur Dokumentation von Argumenten für die Mandantenerklärung auf. Anschließend sollten sie ihre Wahrscheinlichkeitseinschätzung erneut abgeben. Die neue Einschätzung von 75,62% (zuvor 64,62%) wies ebenfalls auf einen signifikanten Einfluss der schriftlichen Begründung hin.

<sup>5</sup> Eine ausführliche Darstellung der Studie erfolgt in Kap. 3.5.5.3.1.2.2.

<sup>6</sup> Vgl. KOONCE/PHILLIPS (1996), S. 36.

tenerklärung auf durchschnittlich 44,1% schätzen. Nachdem dieselben Prüfer die Argumente für die Mandantenerklärung schriftlich niederlegten, korrigierten sie ihre Wahrscheinlichkeitseinschätzung auf durchschnittlich 78,7%.<sup>1</sup>

#### **3.5.5.3.1.2.1.4 Mögliche Debiasing-Mechanismen gegen den written explanation effect**

##### **3.5.5.3.1.2.1.4.1 Dokumentation von Gegenargumenten**

Als einen potenziellen Debiasing-Mechanismus schlägt die psychologische Forschung die Dokumentation von Gegenargumenten vor (sog. *counterexplanation effect*), also die schriftliche Ausführung von Argumenten, die auf die Unrichtigkeit oder Unvollständigkeit der Mandantenerklärung hindeuten.<sup>2</sup> Allerdings lassen Ergebnisse der empirischen Prüfungsforschung vermuten, „that individuals seldom spontaneously counterexplain [...] once a written explanation has been developed.“<sup>3</sup>

Um festzustellen, ob dieser potenzielle debiasing-Mechanismus auch für den Prüfungsbereich in Betracht kommt, führte KOONCE ein zweites Experiment durch, bei welchem die Probanden ebenfalls eine Wahrscheinlichkeitseinschätzung für die Richtigkeit und Vollständigkeit der Mandantenerklärung für die festgestellte Abweichung der Bruttogewinnmarge durchführen sollten.<sup>4</sup> Zuvor sollte ein Teil der Prüfer ausschließlich Argumente aufführen, die für die Richtigkeit und Vollständigkeit der Mandantenerklärung sprechen, wohingegen ein weiterer Teil der Prüfer sowohl Argumente für als auch Argumente gegen die Richtigkeit und Vollständigkeit dokumentieren sollte.<sup>5</sup> Im Unterschied zum ersten Experiment umfassten die Fallunterlagen des zweiten Experiments Informationen, die die Gültigkeit der Mandantenerklärung widerlegen könnten bzw. auf die Gültigkeit anderer Abweichungsursachen hindeuteten.<sup>6</sup> Im Ergebnis stellte KOONCE einen signifikanten Einfluss der zusätzlichen Dokumentation von Gegenargumenten fest. Prüfer, die ausschließlich Argumente für die Mandantenerklärung berücksichtigten, schätzen die Wahrscheinlichkeit der Mandantenerklärung auf 77,2% und Prüfer, die zu-

---

<sup>1</sup> Vgl. KOONCE/PHILLIPS (1996), S. 39.

<sup>2</sup> Vgl. EINHORN/HOGARTH (1986), S. 3-19.

<sup>3</sup> KOONCE (1992), S. 62.

<sup>4</sup> Insgesamt nahmen an diesem Experiment 46 Prüfer mit einer Prüfungserfahrung von 1,9-5,3 Jahren (durchschnittlich 3,2 Jahren) teil.

<sup>5</sup> Vgl. KOONCE (1992), S. 69-71.

<sup>6</sup> Vgl. KOONCE (1992), S. 68-69.



dem Gegenargumente aufführten, auf lediglich 59,65%.<sup>1, 2</sup> Somit scheint die Dokumentation von Gegenargumenten ein potenzielles Mittel darzustellen, um dem *written explanation effect* entgegenzuwirken.

#### 3.5.5.3.1.2.1.4.2 Quantifizierung der Mandantenerklärung

Eine weitere Möglichkeit, der verzerrenden Wirkung des *written explanation effects* entgegenzuwirken, ist in der rechnerischen Überprüfung der Auswirkungen der Mandantenerklärung zu sehen. Ein Modell, in dem dieser Schritt explizit berücksichtigt wird, ist auf ANDERSON/KOONCE zurückzuführen.<sup>3</sup> Sie beschrieben die Bewertung von Hypothesen im Rahmen von analytischen Prüfungshandlungen als einen zweistufigen Prozess. Hierbei wird in einem ersten Schritt zunächst allgemein überprüft, ob der Hypothese überhaupt die Fähigkeit zugesprochen werden kann, die festgestellte Abweichung hervorgerufen zu haben (sog. *plausibility check*). Anschließend wird in einem zweiten Schritt eine Quantifizierung der Mandantenerklärung vorgenommen (sog. *sufficiency check*), die kontrolliert, ob ebenfalls eine betragsmäßige Übereinkunft besteht.

ANDERSON/KOONCE stellten in einer zuvor durchgeführten Studie fest, dass ein signifikanter Zusammenhang zwischen der Annahme einer unvollständigen Mandantenerklärung und der fehlenden Durchführung einer Quantifizierung vorliegt.<sup>4</sup> Um zu untersuchen, ob eine explizite Aufforderung zur Quantifizierung die Chancen erhöht, dass der Prüfer die Unvollständigkeit der Mandantenerklärung erkennt, führten die Autoren eine weitere Studie durch, bei der 36 Prüfer eine wesentliche unerwartete Abweichung in der Bruttogewinnmarge beurteilen sollten. Aus den Fallunterlagen ging hervor, dass die festgestellte Abweichung jedoch nur teilweise (27%) durch die Mandantenerklärung erklärt werden konnte. Ein Teil der Prüfer wurde aufgefordert, ausschließlich eine Plausibilitätskontrolle vorzunehmen, der andere Teil wurde angewiesen, zusätzlich eine Quantifizierung zu praktizieren.<sup>5</sup>

---

<sup>1</sup> Vgl. KOONCE (1992), S. 71-72.

<sup>2</sup> Um zusätzlich beurteilen zu können, ob möglicherweise die Reihenfolge von explanations und counterexplanations einen Einfluss auf die Wahrscheinlichkeitsbeurteilung ausübt, unterteilte KOONCE zusätzlich in Probanden, die zunächst Argumente (Gegenargumente) und anschließend Gegenargumente (Argumente) aufführten. Dabei konnte festgestellt werden, dass die teilnehmenden Prüfer, die zunächst Argumente und anschließend Gegenargumente aufführten, die Wahrscheinlichkeit für die Mandantenerklärung wesentlich niedriger schätzten (53,25%) als die Prüfer, die zunächst Gegenargumente und abschließend Argumente dokumentierten (66,47%). Vgl. KOONCE (1992), S. 72.

<sup>3</sup> Vgl. im Folgenden ANDERSON/KOONCE (1998), S. 3.

<sup>4</sup> Vgl. hierzu ausführlich Kap. 3.5.5.3.1.2.1.2.

<sup>5</sup> Vgl. ANDERSON/KOONCE (1998), S. 4-5.

ANDERSON/KOONCE stellten fest, dass Prüfer mit Quantifizierung signifikant häufiger der Ansicht waren, dass die Mandantenerklärung nicht korrekt sei, als Prüfer, die keine Quantifizierung durchführten. Insgesamt erkannten 16 der 18 Prüfer mit Quantifizierung, dass die Mandantenerklärung nicht vollständig für die festgestellte Abweichung verantwortlich war. Sie beurteilten die Wahrscheinlichkeit, dass die Mandantenerklärung vollständig für die festgestellte Abweichung verantwortlich sei, auf durchschnittlich 25,94%. Hingegen erkannten lediglich sieben der 18 Prüfer ohne Quantifizierung, dass die festgestellte Abweichung nur unvollständig auf die Mandantenerklärung zurückgeführt werden konnte. Sie beurteilten die Wahrscheinlichkeit mit durchschnittlich 54,39%.<sup>1</sup> Somit scheint auch die Quantifizierung einer Mandantenerklärung ein geeignetes Mittel darzustellen, wie einer Verzerrung durch den Erhalt und die Dokumentation einer Mandantenerklärung entgegengewirkt werden kann.

Allerdings ist diesbezüglich einschränkend hinzuzufügen, dass eine weitere Studie von ANDERSON/KADOUS/KOONCE darauf hinweist, dass lediglich eigenständig erarbeitete Kalkulationen einen Einfluss auf die Bewertung einer Mandantenerklärung haben.<sup>2</sup> Sie stellten fest, dass Prüfer die Korrektheit einer Mandantenerklärung vergleichbar hoch einschätzten und zwar unabhängig davon, ob sie zusätzlich zur Erklärung eine Quantifizierung ihrer Auswirkungen vom Mandanten erhielten oder nicht.<sup>3</sup> Die Autoren führten diese Beobachtung darauf zurück, dass die kausalen Zusammenhänge einer Quantifizierung möglicherweise besser nachvollzogen werden können, wenn die Argumentation selbstständig „konstruiert“ wurde.<sup>1</sup>

#### **3.5.5.3.1.2.1.5 Würdigung**

Für die Prüfungspraxis bleibt generell anzumerken, dass durch die Dokumentationspflicht eine potenzielle Gefahr für die Effektivität der Jahresabschlussprüfung ausgeht. Prüfer ordnen inkorrekten bzw. unvollständigen Mandantenerklärungen tendenziell eine höhere Wahrscheinlichkeit zu, wenn sie zuvor die der Mandantenerklärung zugrunde liegende Argumentation schriftlich dokumentiert haben.

Eine Möglichkeit, dieser Verzerrung entgegen zu wirken, ist in der Dokumentation von Gegenargumenten zu sehen. Da sich Prüfer in der Praxis jedoch kaum selbstständig zur

---

<sup>1</sup> Vgl. ANDERSON/KOONCE (1998), S. 8-9.

<sup>2</sup> Eine ausführliche Darstellung des Untersuchungsdesigns findet sich in Kap. 3.5.6.2.3.

<sup>3</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 21-22.

Gegenargumentation motivieren,<sup>2</sup> erscheint es durchaus angemessen, eine derartige Aufforderung in den Prüfungsprozess zu integrieren. Dieses wäre etwa durch die Aufnahme eines entsprechenden Hinweises in einem Prüfungsstandard oder in den Prüfungshandbüchern einfach umzusetzen. Diese Maßnahme könnte durch eine stärkere Betonung der Berücksichtigung von Gegenargumenten im Rahmen von Schulungsmaßnahmen sinnvoll flankiert werden.

Weiterhin konnte gezeigt werden, dass auch die Aufforderung zur Vornahme einer Quantifizierung einen positiven Einfluss auf eine angemessene Bewertung einer unvollständigen Mandantenerklärung ausüben kann. Um auch von diesem debiasing-Mechanismus profitieren zu können, sollten die Prüfungsgesellschaften möglicherweise in Betracht ziehen, ihre Mitarbeiter auf die Notwendigkeit einer rechnerischen Kontrolle hinzuweisen. Es erscheint durchaus sinnvoll, diese Maßnahme in Kombination mit der o.a. Aufforderung zur Gegenargumentation umzusetzen.

#### **3.5.5.3.1.2.2 Der Einfluss der Verständlichkeit einer Mandantenerklärung**

In der psychologischen Forschung konnte nachgewiesen werden, dass Wahrscheinlichkeitsbeurteilungen durch den Grad der Verständlichkeit einer Erklärung beeinflusst werden können.<sup>3</sup> So beurteilten Individuen die Wahrscheinlichkeit einer Erklärung als sehr hoch, wenn ihnen die dazu notwendigen Informationen leicht zugänglich waren und sie das zusammenhängende Muster zügig erschließen konnten. Umgekehrt schätzen sie die Wahrscheinlichkeit einer Ursache als gering ein, wenn diese schwer verständlich war. Darüber hinaus konnte ebenfalls festgestellt werden, dass die Leichtigkeit bzw. Schwierigkeit, mit der eine Erklärung verstanden wurde, zudem die Beurteilung alternativer Erklärungen beeinflusste. Lag etwa eine einfach zu erschließende Erklärung vor, so wurden andere Erklärungen nicht weiter in Betracht gezogen, da die erste Erklärung möglicherweise als ausreichend angesehen wurde.<sup>4</sup>

Im Rahmen analytischer Prüfungshandlungen könnte eine derartige Beeinflussbarkeit des Prüfers schwerwiegende Folgen mit sich bringen. Die Mitarbeiter des zu prüfenden Unternehmens könnten etwa bemüht sein, Erklärungen für eine wesentliche Abwei-

---

<sup>1</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 25.

<sup>2</sup> So auch KOONCE (1992), S. 74-75.

<sup>3</sup> Vgl. PENNINGTON/HASTIE (1988), S. 521-533 sowie SHERMAN/CIALDINI/SCHWARTZMAN/REYNOLDS (1985), S. 118-127.

<sup>4</sup> Vgl. EINHORN/HOGARTH (1986), S. 3-19.

chung möglichst verständlich zu begründen. Da generell nicht auszuschließen ist, dass Fragen zu Diskrepanzen im Jahresabschluss von Mandanten antizipiert werden können, stünde ihnen sogar ausreichend Zeit zur Verfügung, um entsprechende Informationen zusammenzutragen.<sup>1</sup>

Um zu untersuchen, ob eine derartige Verzerrung der Hypothesenbeurteilung auch im Kontext der Abschlussprüfung vorliegt, führten KOONCE/PHILLIPS eine Studie durch, bei der sie 38 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 2,9 Jahren eine wesentliche Abweichung in der Bruttogewinnmarge analysieren ließen. Dabei wurde allen Prüfern eine inkorrekte Mandantenerklärung sowie sechs weitere potenzielle Abweichungsursachen zur Verfügung gestellt. Der Mandant führte die Abweichung auf die Änderung des Verkaufssortiments zurück. Den Fallunterlagen konnten Informationen entnommen werden, die auf eine mögliche Korrektheit der Mandantenerklärung hinwiesen. Die tatsächliche Abweichungsursache konnte jedoch nicht eindeutig identifiziert werden.<sup>2</sup> Die Prüfer wurden in zwei Gruppen unterteilt. Bei einer Gruppe wurden die Informationen zur Beurteilung der Mandantenerklärung leicht verständlich aufbereitet, so dass der Prüfer den Prozess, wie die Änderung des Sortiments zur festgestellten Abweichung geführt haben könnte, einfach nachvollziehen konnte. Bei der anderen Prüfergruppe wurden die gleichen Informationen unverständlich dargeboten, so dass die Argumentation nur schwierig nachvollzogen werden konnte. Anschließend wurden alle Prüfer gefragt, wie hoch sie die Wahrscheinlichkeit einschätzen, dass (1) die Mandantenerklärung und (2) eine der anderen sechs Hypothesen der tatsächliche Grund der Abweichung ist.<sup>3</sup>

Die Autoren stellten fest, dass die Prüfergruppe mit den einfach nachzuvollziehenden Informationen die Mandantenerklärung signifikant wahrscheinlicher (62,08%) einschätzten als die Prüfer mit den schwierig nachzuvollziehenden Informationen (31,11%). Allerdings beurteilte die erste Prüfergruppe die Wahrscheinlichkeit der anderen Hypothesen nicht signifikant niedriger als die zweite Prüfergruppe (41,26% ggü. 46,16%).<sup>1</sup>

Für die Prüfungspraxis sind diese Erkenntnisse von enormer Bedeutung, da sie darauf hinweisen, dass die Mandantenbefragung zu einer erheblichen Beeinflussung der Ver-

---

<sup>1</sup> Vgl. etwa KOONCE/PHILLIPS (1996), S. 33-34. Ähnlich argumentieren auch ANDERSON/KOONCE/MARCHANT (1994), S. 138.

<sup>2</sup> Vgl. ANDERSON/PHILLIPS (1996), S. 35-36.

<sup>3</sup> Vgl. KOONCE/PHILLIPS (1996), S. 37.

ständigkeit und Bewertung von Abweichungshypothesen führen kann. Sollte der Prüfer eine inkorrekte Mandantenerklärung aufgrund der hohen Verständlichkeit der Argumentationsweise des Mandanten als wahrscheinlich einschätzen, so könnte dieses die Effektivität von analytischen Prüfungshandlungen erheblich beeinträchtigen. Andererseits könnte eine geringe Wahrscheinlichkeitsbeurteilung einer korrekten aber schwierig nachzuvollziehenden Mandantenerklärung zu Effizienzeinbußen führen, da weitere auf die Beurteilung anderer Abweichungsursachen bezogene Prüfungshandlungen zu höheren Prüfungskosten führen könnten.<sup>2</sup>

### **3.5.5.3.1.2.3 Der Einfluss der Kompetenz der Informationsquelle**

In einem engen Zusammenhang zur Verständlichkeit einer Erklärung für eine festgestellte Abweichung steht die Kompetenz des Erklärenden. Als Kompetenz bezeichnet man im Allgemeinen „the technical ability of the individual providing the information“<sup>3</sup> und zwar unabhängig von der Korrektheit der Informationen.<sup>4</sup> Besitzt demnach ein Individuum die Fähigkeit, die für eine Erklärung notwendigen Informationen technisch in einer verständlichen Art und Weise zu vermitteln, ist dieser Person eine hohe Kompetenz zuzuordnen.

Indes weisen psychologische Theorien darauf hin, dass die Überzeugungskraft eines Individuums in Abhängigkeit von seiner Kompetenz zu betrachten ist.<sup>5</sup> Demnach ist einer kompetenten Person unabhängig von der Richtigkeit der Erklärung eine höhere Überzeugungsfähigkeit zuzusprechen als Personen mit geringerer Kompetenz. Im Kontext der Mandantenbefragung zur Identifikation einer unerwarteten Schwankung im Jahresabschluss könnte dies eine Beeinträchtigung der Prüfungsqualität nach sich ziehen, da vermutet werden kann, dass der Mandant seine möglicherweise inkorrekte bzw. unvollständige Abweichungsbegründung möglichst überzeugend vorträgt, um den Prüfer für die Akzeptanz seiner Erklärung zu gewinnen.<sup>6</sup> Gelingt es ihm dabei, eine erhöhte Kom-

---

<sup>1</sup> Vgl. KOONCE/PHILLIPS (1996), S. 39-40.

<sup>2</sup> Vgl. KOONCE/PHILLIPS (1996), S. 43.

<sup>3</sup> REIMERS/FENNEMA (1999), S. 117.

<sup>4</sup> Vgl. SCHREIBER (2000), S. 143.

<sup>5</sup> Vgl. EAGLY/CHAIKEN (1984), S. 267-359.

<sup>6</sup> Vgl. ANDERSON/KOONCE/MARCHANT (1994), S. 138.

petenz „vorzutäuschen“, so könnte sich dies verzerrend auf die Beurteilung der mitgeteilten Abweichungsursache auswirken.<sup>1</sup>

In diesem Zusammenhang ist die Studie von ANDERSON/KOONCE/MARCHANT zu betrachten. Sie ließen 124 Prüfer mit einer durchschnittlichen Prüfungserfahrung vom 74 Monaten eine wesentliche Abweichung in der Bruttogewinnmarge analysieren. Allen Teilnehmern wurde mitgeteilt, dass der Controller die Schwankung auf eine Änderung des Verkaufssortiments zurückführt. Allerdings erhielten die Probanden unterschiedliche Hinweise bezüglich der Kompetenz des Controllers. Bei der einen Gruppe wurde der Controller als besonders kompetent beschrieben. Die andere Gruppe erhielt hingegen die Information, dass dem Controller eine geringe Kompetenz zu attestieren ist. Abschließend wurden alle Teilnehmer gefragt, wie hoch sie die Wahrscheinlichkeit für die Richtigkeit der Mandantenerklärung einschätzen.<sup>2</sup>

Dabei stelle sich heraus, dass die Kompetenz des Controllers einen signifikanten Einfluss auf die Beurteilung der mitgeteilten Hypothese ausübt. Prüfer, welche die Mandantenerklärung von einem kompetenten Controller erhielten, beurteilten diese mit einer Wahrscheinlichkeit von 56,28% im Vergleich zu 46,39% bei Prüfern, denen die Abweichungsursache von einem weniger kompetenten Controller zur Verfügung gestellt wurde.<sup>3, 4</sup>

---

<sup>1</sup> Ähnlich argumentiert HIRST (1994), S. 116.

<sup>2</sup> Vgl. ANDERSON/KOONCE/MARCHANT (1994), S. 140.

<sup>3</sup> Vgl. ANDERSON/KOONCE/MARCHANT (1994), S. 141.

<sup>4</sup> In einer ähnlichen Studie von HIRST (1994), die ebenfalls den Zusammenhang zwischen der Kompetenz einer Quelle und der Beurteilung der Richtigkeit ihrer Erklärung untersuchte, konnte dieser Einfluss hingegen nicht nachgewiesen werden. Allerdings kann diese Studie nicht direkt mit der Studie von ANDERSON/KOONCE/MARCHANT verglichen werden, da sich beide Studien in einem wesentlichen Punkt voneinander differenzieren. Bei HIRST wurde den Prüfern mitgeteilt, dass es sich bei der Informationsquelle um den Finanzvorstand (CFO) handelte. Bei ANDERSON/KOONCE/MARCHANT hingegen handelte es sich um einen Assistenten der Controlling-Abteilung. Während es durchaus denkbar erscheint, dass einem neu eingestellten Assistenten, der zudem kaum Erfahrung im Bereich der Rechnungslegung mitbringt, wesentlich weniger Kompetenz zugeschrieben wird als einem Assistenten mit umfangreicher Erfahrung im Umgang mit Rechnungslegungsfragen (so geschehen bei ANDERSON/KOONCE/MARCHANT (1994), S. 140), erscheint es durchaus möglich, dass einem Finanzvorstand in Rechnungslegungsangelegenheiten generell eine hohe Kompetenz zugeordnet wird und zwar unabhängig davon, in welcher Branche er zuvor tätig war (so geschehen bei HIRST (1994), S. 119). Insofern ist es möglich, dass die Probanden bei HIRST, denen die Abweichungsursache von einem „nicht kompetenten“ Finanzvorstand vorgetragen wurde, diesen zumindest in grundlegenden Rechnungslegungsfragen für ausreichend kompetent hielten. Dabei ist auch zu berücksichtigen, dass es sich um eine Schwankung in einer branchenübergreifenden bzw. elementaren Kennzahl handelte. Im Vergleich zu einem Finanzvorstand ist es bei einem berufsunerfahrenen Assistenten durchaus denkbar, dass es ihm an elementaren Rechnungslegungskennnissen mangelt und mithin eine geringe Kompetenz zugeordnet wird.

#### 3.5.5.3.1.2.4 Einfluss der Integrität bzw. Objektivität der Informationsquelle

HIRST beurteilte die Vorgehensweise von ANDERSON/KOONCE/MARCHANT als kritisch. Er verweist darauf, dass die alleinige Betrachtung der Kompetenz nicht ausreiche, da Kompetenz nicht zwangsläufig bedeute, dass wahrheitsgemäß berichtet wird. Vielmehr sei neben der Kompetenz einer Quelle auch ihre Integrität (bzw. Objektivität) zu berücksichtigen.<sup>1</sup> Integrität „can be defined as the likelihood that an individual will report the information truthfully.“<sup>2</sup>

Eine Studie, die u.a. den Einfluss der Integrität des Mandanten auf die Beurteilung einer von diesem abgegebenen Abweichungserklärung untersucht, führte PEECHER durch.<sup>3</sup> Er ließ 106 Prüfer mit einer durchschnittlichen Prüfungserfahrung von drei Jahren eine analytische Prüfungshandlung durchführen, bei der eine wesentliche Abweichung in den Verbindlichkeiten einerseits und der Bruttogewinnmarge andererseits zu beurteilen waren. Den Probanden wurde mitgeteilt, dass der Mandant diese Schwankungen auf eine Verlängerung des Zahlungsziels gegenüber den Kunden bzw. eine Änderung des Verkaufssortiments zurückführte.<sup>4</sup> PEECHER ließ dabei einer Gruppe der Prüfer den Hinweis zukommen, dass dem Mandanten eine geringe Integrität zuzuordnen sei. Die anderen Prüfer informierte er hingegen darüber, dass dem Mandanten eine höhere Integrität beizumessen sei. Anschließend wurden alle Probanden aufgefordert, eine Wahrscheinlichkeitsbeurteilung hinsichtlich der Korrektheit der Mandantenerklärung abzugeben.<sup>5</sup>

Insgesamt wiesen die Ergebnisse darauf hin, dass Prüfer die Korrektheit einer Abweichungshypothese im Schnitt signifikant wahrscheinlicher einschätzten, wenn sie von ei-

---

<sup>1</sup> Vgl. HIRST (1994), S. 115-116.

<sup>2</sup> REIMERS/FENNEMA (1999), S. 117. Eine ähnliche Definition findet sich bei HIRST (1994), S. 116.

<sup>3</sup> Im Rahmen von analytischen Prüfungshandlungen untersuchte auch HIRST (1994) u.a. den Einfluss der Integrität einer Informationsquelle auf die Hypothesenbewertung. Allerdings differenzierte HIRST zwischen einer Erklärung eines Mitglieds des Prüfungsteams und eines Mandanten, wobei Erstgenanntem eine höhere und Letztgenanntem eine geringere Integrität beigemessen wurde. Dabei konnte festgestellt werden, dass Prüfer einer Aussage eines Mitglieds des Prüfungsteams eine höhere Eintrittswahrscheinlichkeit zuordneten als der Aussage eines Mitarbeiters des Mandanten. Zudem stellte HIRST fest, dass die Eintrittswahrscheinlichkeit einer Aussage eines Prüfungsteammitglieds signifikant höher eingeschätzt wurde, wenn dieser eine höhere Kompetenz aufweist. Diese Erkenntnisse weisen darauf hin, dass die Vorgehensweise der Prüfer in Einklang mit den berufsrechtlichen Normen steht, da ISA 520.12c-d vorschreibt, der Information einer externen Quelle eine höhere Zuverlässigkeit beizumessen als der Information einer internen Quelle des geprüften Unternehmens.

<sup>4</sup> Vgl. PEECHER (1996), S. 128 i.V.m. 130.

<sup>5</sup> Vgl. PEECHER (1996), S. 129.

nem Mandanten mit höherer Integrität geäußert wurde (63%) im Vergleich zu einem Mandanten mit geringerer Integrität (53%).<sup>1, 2</sup>

### **3.5.5.3.1.3 Der Einfluss des Erhalts einer Mandantenerklärung auf den (Gesamt-) Erfolg einer analytischen Prüfungshandlung**

Obwohl in dem vorliegenden Kap. bereits eindeutige empirische Belege für den verzerrenden Einfluss einer Mandantenerklärung auf die Hypothesenbewertung erbracht wurden, konnte bisher noch nicht nachgewiesen werden, dass Prüfer ohne Mandantenerklärung generell erfolgreicher bei der Identifikation einer Abweichungsursache sind. In diesem Kontext ist die Untersuchung von GREEN zu sehen.<sup>3</sup> Sie unterteilte ihre Probanden in zwei Gruppen, wobei der einen Gruppe eine inkorrekte Mandantenerklärung mitgeteilt wurde, der zweiten Gruppe hingegen nicht. Im Ergebnis stellte GREEN überraschenderweise fest, dass beide Gruppen eine vergleichbare Erfolgsquote aufwiesen. Bei den Prüfern mit Mandantenerklärung waren lediglich 24% in der Lage, den tatsächlichen Grund der Abweichung zu identifizieren. Weitere 40% wählten die inkorrekte Mandantenerklärung als Abweichungsursache aus, während weitere 33% einen anderen Nicht-Fehlergrund und die übrigen Prüfer (2%) einen anderen Fehlergrund selektierten.<sup>1</sup> Demgegenüber waren die Prüfer ohne Mandantenerklärung nicht erfolgreicher. Von ihnen waren lediglich 25% befähigt, den richtigen Grund festzustellen. Die weiteren 75% wählten einen selbstgenerierten Nicht-Fehlergrund, wobei keiner identisch mit der Mandantenerklärung war.<sup>2</sup>

Insofern kann insgesamt konstatiert werden, dass der Erhalt einer Mandantenerklärung zwar regelmäßig als Grund für die festgestellte Abweichung gewählt wird. Allerdings sind Prüfer, die keine Mandantenerklärung erhalten haben, nicht eher dazu in der Lage, die richtige Ursache zu identifizieren. Sie wählen zwar nicht die Mandantenerklärung aus, aber dafür eine andere inkorrekte Abweichungsursache.

---

<sup>1</sup> Vgl. PEECHER (1996), S. 134.

<sup>2</sup> Auch KAPLAN/RECKERS (1989), S. 544 konnten feststellen, dass die Integrität des Mandanten einen wesentlichen Einfluss auf die Hypothesenbeurteilung ausübt. Die Autoren ließen ihren Probanden zwar keine explizite Mandantenerklärung zukommen. Allerdings konnten sie im Rahmen ihrer Studie nachweisen, dass Prüfer die Wahrscheinlichkeit einer Fehlerursache signifikant höher (geringer) einschätzen, wenn der Mandant durch eine geringe (hohe) Integrität geprägt war.

<sup>3</sup> Eine detaillierte Darstellung der Studie erfolgte bereits in Kap. 3.5.3.3.2.1.2.1.



### **3.5.5.3.2 Der Erhalt einer Abweichungshypothese durch Entscheidungshilfen**

#### **3.5.5.3.2.1 Vorbemerkungen**

Softwaregestützte Entscheidungshilfen können ein probates Mittel darstellen, um sicherzustellen, dass der Prüfer bei der Durchführung von analytischen Prüfungshandlungen eine ausreichende Anzahl an potenziellen Ursachen berücksichtigt, die der identifizierten Abweichung zugrunde liegen könnten.<sup>3</sup> Dieses scheint insbesondere vor dem Hintergrund sinnvoll, dass Prüfer häufig Schwierigkeiten haben, eine ausreichende Anzahl an möglichen Abweichungsursachen eigenständig zu generieren.<sup>4</sup>

Neben diesem potenziellen Nutzen von Entscheidungshilfen sollte jedoch auch in Betracht gezogen werden, dass die Verwendung von Entscheidungshilfen möglicherweise eine Gefährdung der prüferischen Vorgehensweise nach sich zieht. So konnte bereits mehrfach nachgewiesen werden, dass sich Entscheidungshilfen negativ auf die Problemlösung auswirken, etwa weil sich Individuen unverhältnismäßig hoch auf die Entscheidungshilfe verlassen und ihre Aufmerksamkeit entsprechend einschränken.<sup>5</sup> Dieses könnte u.a. dazu führen, dass Nutzer von Entscheidungshilfen wichtigen Hinweisen eine zu geringe Aufmerksamkeit widmen und ihren kognitiven Aufwand minimieren. BAMBER ET AL. wiesen daher frühzeitig auf die Notwendigkeit hin, dass vor der Entwicklung und Implementierung von Entscheidungshilfen zunächst ein umfassendes Wissen zu potenziellen Auswirkungen auf die prüferische Vorgehensweise erlangt werden sollte.<sup>6</sup>

#### **3.5.5.3.2.2 Der Einfluss einer Fehler- bzw. Nicht-Fehler-Dominanz von Entscheidungshilfen auf die Hypothesenbewertung**

Um festzustellen, ob eine fehler- bzw. nichtfehlerdominierte Entscheidungshilfe einen Einfluss auf die Hypothesenbewertung hat, ließen ANDERSON/KAPLAN/RECKERS 184 Prüfer mit einer durchschnittlichen Prüfungserfahrung von etwa vier Jahren eine 5%-ige

---

<sup>1</sup> Vgl. GREEN (2004a), S. 386.

<sup>2</sup> Vgl. GREEN (2004a), S. 386.

<sup>3</sup> Zur Anwendung von Entscheidungshilfen siehe insbesondere auch Kap. 3.5.5.3.2.1.

<sup>4</sup> HEIMAN (1990), S. 883 etwa stellte in ihrer Untersuchung fest, dass es elf von 35 Prüfern nicht möglich war, mehr als zwei alternative Fehlerhypothesen für eine wesentliche Abweichung im Jahresabschluss zu generieren. Dass Prüfer Schwierigkeiten bei der Generierung einer Mehrzahl von Abweichungshypothesen aufweisen, konnten auch ANDERSON/KAPLAN/RECKERS (1992), S. 10 nachweisen.

<sup>5</sup> Vgl. stellvertretend PINCUS (1989), S. 153-163 sowie KACHELMEIER/MESSIER (1990), S. 209-226.

<sup>6</sup> Vgl. BAMBER/GILLET/MOCK/TROTMAN (1995), S. 55-85.

Abweichung der Umschlagshäufigkeit der Vorräte untersuchen.<sup>1</sup> Die Probanden wurden zunächst aufgefordert, die Wahrscheinlichkeit einzuschätzen, dass es sich bei der zugrunde liegenden Ursache um einen Fehler bzw. Nicht-Fehler handelt. Anschließend wurde den Prüfern eine Liste mit 15 potenziellen Abweichungsursachen zur Verfügung gestellt.<sup>2</sup> Dabei unterteilten die Autoren ihre Teilnehmer in zwei Gruppen. Der einen Hälfte der Prüfer wurde eine fehlerdominierte Liste und der anderen Hälfte eine nicht-fehlerdominierte Liste präsentiert. In Abhängigkeit der Dominanz enthielt die Liste elf Fehler- bzw. Nicht-Fehlerursachen und lediglich vier Nicht-Fehler- bzw. Fehlerursachen. Nachdem die Prüfer die vorgeschlagenen potenziellen Abweichungsursachen zur Kenntnis genommen haben, wurden sie erneut gebeten, die Wahrscheinlichkeit eines Fehlers bzw. Nicht-Fehlers einzuschätzen.<sup>3</sup>

Im Ergebnis konnten die Autoren feststellen, dass die Bereitstellung einer fehlerdominierten Liste einen signifikanten Einfluss auf die Wahrscheinlichkeitsbeurteilung bewirkte. Prüfer in dieser Gruppe erhöhten ihre Fehlerwahrscheinlichkeit von ursprünglich 32,7% auf 38,9%. Umgekehrt konnte kein signifikanter Einfluss einer nicht-fehlerdominierten Liste festgestellt werden (33,6% im Vgl. zu 30,9%).<sup>4, 5</sup>

Prüfungsgesellschaften, die sich überwiegend um eine hohe Effektivität von Jahresabschlussprüfungen bemühen, könnten diese Ergebnisse zum Anlass nehmen, um auf fehlerdominierte Entscheidungshilfen zurückzugreifen, welche die Aufmerksamkeit des Prüfers auf die Identifikation von Fehlerursachen fokussieren.<sup>6</sup> Aus Effizienzgründen wäre der Sachverhalt allerdings genau umgekehrt zu beurteilen.

---

<sup>1</sup> Vgl. ANDERSON/KAPLAN/RECKERS (1997), S. 6-7.

<sup>2</sup> Dieses sollte die Verwendung einer Entscheidungshilfe simulieren.

<sup>3</sup> Vgl. ANDERSON/KAPLAN/RECKERS (1997), S. 7.

<sup>4</sup> Vgl. ANDERSON/KAPLAN/RECKERS (1997), S. 10, 12.

<sup>5</sup> Weiterhin wurde kein signifikanter Einfluss von Erfahrung festgestellt. „Regarding experience, for the error dominated list, managing auditors were just as likely as the less experience audit seniors to increase their assessment of the probability of error, when presented with a list that was primarily error based.” (ANDERSON/KAPLAN/RECKERS (1997), S. 13). Da die Autoren zusätzlich das Prüfungsrisiko variierten, konnte zudem festgestellt werden, dass die Wahrscheinlichkeitsbeurteilung eines Fehlers wesentlich durch die Risikoeinschätzung beeinflusst wurde (22,9% bei geringem Risiko im Vergleich zu 46,1% bei hohem Risiko). Kritisch zu bewerten ist allerdings, dass eine fehlerdominierte Liste auch bei niedrigem Prüfungsrisiko zu einer signifikanten Erhöhung der Fehlerwahrscheinlichkeit führte, was sich möglicherweise negativ auf die Effizienz von Abschlussprüfungen auswirken könnte. Vgl. ANDERSON/KAPLAN/RECKERS (1997), S. 12.

<sup>6</sup> Vgl. HERMANSON (1997), S. 21.

### **3.5.5.3.2.3 Der Einfluss der Verlässlichkeitseinschätzung auf die Hypothesenbewertung**

Die Beurteilung der Verlässlichkeit einer Informationsquelle spielt eine wesentliche Rolle im Rahmen der Durchführung von analytischen Prüfungshandlungen. Dieses wird durch die ISA 520.12c (a) dokumentiert, der erst kürzlich um einen Hinweis erweitert wurde, der zur Beurteilung der Beweiskraft einer Information auf die Berücksichtigung der Objektivität der jeweiligen Quelle verweist.

Da Entscheidungshilfen möglicherweise als eine objektivere Quelle als etwa der Mandant oder dessen Mitarbeiter eingeschätzt werden, könnten Prüfer tendenziell dazu neigen, einer Abweichungserklärung, die von einer Entscheidungshilfe zur Verfügung gestellt wird, eine höhere Eintrittswahrscheinlichkeit zuzuordnen als derselben Abweichungserklärung, wenn sie vom Mandanten mitgeteilt wird.<sup>1</sup>

In diesem Kontext führten ANDERSON/MORENO/MUELLER eine Studie durch, in der sie 51 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 74 Monaten eine wesentliche Abweichung in der Umschlagshäufigkeit der Vorräte untersuchen ließen.<sup>2</sup> Dabei unterteilten sie die Prüfer in zwei Gruppen. Eine Prüfergruppe erhielt insgesamt drei plausible Nicht-Fehlerursachen von einer Entscheidungshilfe. Der anderen Gruppe wurden die gleichen Hypothesen „vererbt“, allerdings nahmen diese Prüfer die Hypothesen von dem Mandanten entgegen. Den Prüfern war es unabhängig von der Informationsquelle möglich, die Unvollständigkeit der drei Hypothesen anhand der weiteren Fallunterlagen zu erkennen, da sie lediglich 50% der festgestellten Abweichung verursachten. Abschließend wurden alle Prüfer aufgefordert, die Wahrscheinlichkeit zu beurteilen, dass die festgestellte Abweichung vollständig auf die drei mitgeteilten Hypothesen zurückgeführt werden kann.<sup>3, 4</sup>

Im Ergebnis konnten die Autoren feststellen, dass die Prüfer die potenziellen Abweichungsursachen signifikant wahrscheinlicher einschätzten, wenn sie diese von einer Entscheidungshilfe erhielten. Im Schnitt beurteilten diese Prüfer die Wahrscheinlichkeit

---

<sup>1</sup> Vgl. ANDERSON/MORENO/MUELLER (2003), S. 2.

<sup>2</sup> Vgl. ANDERSON/MORENO/MUELLER (2003), S. 4-5.

<sup>3</sup> Vgl. ANDERSON/MORENO/MUELLER (2003), S. 5-6.

<sup>4</sup> Um sicherzustellen, dass der Einfluss nicht auf Unterschiede in der Kompetenzbeurteilung hinsichtlich des Mandanten bzw. der Entscheidungshilfe zurückgeführt werden kann, wurden die Fallunterlagen derart konzipiert, dass beiden Quellen eine vergleichbare Kompetenz zugeordnet werden konnte. Anhand einer post-experimentellen Befragung der Teilnehmer konnte bestätigt werden, dass keine signifikanten Differenzen in Bezug auf die Kompetenzbeurteilung der beiden Quellen vorlag. Vgl. ANDERSON/MORENO/MUELLER (2003), S. 6-7.

auf einer Likert-Skala von eins bis elf mit 7,37 Punkten. Prüfer, die die identischen Hypothesen vom Mandanten entgegennahmen, bewerteten diese durchschnittlich mit 5,79 Punkten.<sup>1</sup> Die Autoren führten diese signifikanten Unterschiede in den Wahrscheinlichkeitsurteilen auf die unterschiedlichen Objektivitätseinschätzungen gegenüber einer Entscheidungshilfe einerseits und einem Mandanten andererseits zurück. Erstgenannte wurde von den entsprechenden Teilnehmern mit einer durchschnittlichen Punktezahl von 7,17 bewertet, Letztgenannte hingegen signifikant geringer mit durchschnittlich 2,30 Punkten.<sup>2</sup>

Diese Erkenntnisse zeigen, dass Prüfer tendenziell dazu geneigt sein könnten, die Objektivität einer Entscheidungshilfe äußerst hoch einzuschätzen und dieses zu einer Annahme von vorgeschlagenen Abweichungsursachen führen könnte, obwohl die Abweichung nicht oder nur teilweise auf die „vererbten“ Hypothesen zurückgeführt werden kann. Für die Prüfungspraxis wird daher empfohlen, Prüfer in Umgang mit Entscheidungshilfen zu schulen. Insbesondere sollten Prüfer angehalten werden, Abweichungsursachen auch dann skeptisch zu betrachten, wenn diese von einer höchst objektiven Quelle, wie etwa der Entscheidungshilfe, zur Verfügung gestellt werden.<sup>3</sup>

#### **3.5.5.4 Die Bewertung einer Vielzahl von Hypothesen**

##### **3.5.5.4.1 Vorbemerkung**

In der empirischen Prüfungsforschung konnte bereits mehrfach nachgewiesen werden, dass Prüfer eine Vielzahl von Hypothesen bilden, um den Grund einer festgestellten Abweichung identifizieren zu können.<sup>4</sup> Unabhängig davon, ob es sich dabei um eine interne oder externe Hypothese handelt, stellt sich generell die Frage, wie Prüfer eine Vielzahl von unterschiedlichen Hypothesen bewerten.

Im Bereich der empirischen Prüfungsforschung wurden im Zusammenhang mit der multiplen Hypothesenbewertung zwei Aspekte untersucht, die den Inhalt der beiden folgenden Abschnitte einnehmen. Dabei handelt es sich einerseits um das eingesetzte Verfahren zur Beurteilung multipler Hypothesen und andererseits um die Einhaltung bzw. Verletzung der Konjunktionsregel bei der Bewertung mehrerer Hypothesen.

---

<sup>1</sup> Vgl. ANDERSON/MORENO/MUELLER (2003), S. 7.

<sup>2</sup> Vgl. ANDERSON/MORENO/MUELLER (2003), S. 6-7.

<sup>3</sup> So auch ANDERSON/MORENO/MUELLER (2003), S. 8.

<sup>4</sup> Siehe hierzu etwa Kap. 3.5.3.3.1.3.2.

#### 3.5.5.4.2 Verfahren zur Beurteilung einer Vielzahl von Hypothesen

Psychologische Erkenntnisse weisen darauf hin, dass Individuen zur Bewertung multipler Hypothesen zwei unterschiedliche Verfahren zur Verfügung stehen. Hierbei handelt es sich einerseits um das komplementäre Verfahren und andererseits um das unabhängige Verfahren.

Das *komplementäre Verfahren* beruht auf der Annahme, dass alle generierten Hypothesen der gleichen Grundgesamtheit angehören und allen Hypothesen eine bestimmte anfängliche Eintrittswahrscheinlichkeit zuzuordnen ist, die sich in der Summe zu einer Gesamtwahrscheinlichkeit von 100% addieren lassen.<sup>1</sup> Werden im Anschluss weitere Erkenntnisse erlangt, die die Wahrscheinlichkeit einer Hypothese erhöhen (reduzieren), so sind zwangsläufig auch die Wahrscheinlichkeitsurteile hinsichtlich der anderen Hypothesen anzupassen.

Das *unabhängige Verfahren* hingegen betrachtet die jeweiligen Hypothesen als einzelne Einheiten, die nicht zur gleichen Grundgesamtheit gehören.<sup>2</sup> Demnach führen neu erlangte Erkenntnisse lediglich zur Änderung der Wahrscheinlichkeit der zugrunde liegenden Hypothese, nicht jedoch zur Anpassung der Eintrittswahrscheinlichkeiten der anderen Hypothesen. Da die Hypothesen unabhängig voneinander zu beurteilen sind, beläuft sich die Gesamtwahrscheinlichkeit der Hypothesen nicht auf einen festen Wert. Vielmehr kann sich die Gesamtwahrscheinlichkeit auf mehr bzw. weniger als 100% belaufen.<sup>3</sup>

Welches Verfahren in der Prüfungspraxis angewandt wird, kann für die Effizienz von analytischen Prüfungshandlungen von entscheidender Bedeutung sein. So kann etwa durch die Verwendung des komplementären Verfahrens, welches die Hypothesen als zueinander gehörende Einheiten betrachtet, die volle Beweiskraft von Prüfungsnachweisen zum Tragen kommen, da Prüfungsnachweise eine duale Funktion erfüllen. Die direkte Bestätigung einer Hypothese bedeutet gleichzeitig die indirekte Widerlegung der anderen Hypothese. Der Prüfer ist demnach nicht darauf angewiesen, seine Hypothesen direkt zu prüfen. Vielmehr kann er die Wahrscheinlichkeit einer alternativen Hypothese

---

<sup>1</sup> Vgl. GAVANSKI/HUI (1992), S. 766-780 sowie SHERMAN/MCMULLEN/GAVANSKI (1992), S. 401-421; TIEGEN (1983), S. 97-105; TIEGEN (1988), S. 157-174.

<sup>2</sup> Vgl. ROBINSON/HASTIE (1985), S. 443-456.

<sup>3</sup> Vgl. VAN WALLENDael/HASTIE (1990), S. 240-250.

erhöhen (verringern), wenn er aufgrund neuer Informationen die geprüfte Hypothese als (un-)wahrscheinlicher erachtet als zuvor.<sup>1</sup>

Im Gegensatz hierzu werden Prüfungsnachweise bei dem unabhängigen Verfahren immer nur zur Bewertung einer Hypothese herangezogen. Dieses sog. „one hypothesis“-Syndrom führt regelmäßig dazu, dass eine Änderung der Eintrittswahrscheinlichkeit einer Hypothese keine Anpassung der Wahrscheinlichkeiten der weiteren Hypothesen mit sich bringt.<sup>2</sup> Das vollständige Potenzial von Prüfungsnachweisen wird hier also nicht ausgeschöpft.<sup>3</sup>

Obwohl davon auszugehen ist, dass der in den letzten Jahren zunehmende Wettbewerbsdruck auf dem Prüfungsmarkt zu einer erhöhten Notwendigkeit kostengünstiger Prüfungshandlungen geführt haben dürfte, sprechen dennoch zwei Argumente gegen die Verwendung eines komplementären Verfahrens. Erstens, da Prüfer verantwortlich gegenüber Dritten sind (z.B. dem Mandanten, der WPK oder Gerichten) und ihre Vorgehensweise möglicherweise rechtfertigen müssen, könnten Prüfer abgeneigt sein, eine Hypothese aufgrund indirekter Nachweise zu akzeptieren bzw. auszuschließen. Der Prüfer könnte sich durch die Erlangung direkter Nachweise indes sicherer fühlen.<sup>4</sup> Zweitens, das komplementäre Verfahren setzt implizit eine uneingeschränkte Rationalität des Prüfers voraus. Da Prüfer jedoch gewissen kognitiven Beschränkungen<sup>5</sup> unterliegen, tendieren sie möglicherweise zur Verwendung von Mechanismen, die eine Reduzierung der Komplexität erlauben. Im Kontext der Bewertung von multiplen Hypothesen würde dies für die Benutzung eines unabhängigen Verfahrens sprechen.<sup>6</sup>

In Bezug auf die Durchführung von analytischen Prüfungshandlungen beschäftigte sich erstmalig HEIMAN mit den angewandten Bewertungsverfahren von Prüfern. Sie ließ 73 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,5 Jahren eine wesentliche Abweichung in der Bruttogewinnmarge untersuchen. Dabei sollten sie angeben, wie hoch sie die Wahrscheinlichkeit einschätzten, dass diese Abweichung auf eine fehlerhafte Periodenabgrenzung im Bereich der Umsätze (Zielhypothese) zurückzuführen ist. Anschließend generierten die Prüfer alternative Hypothesen, die die festgestellte Abweichung verursacht haben könnten, um abschließend eine erneute Bewertung der Ziel-

---

<sup>1</sup> Vgl. ASARE/RIGHT (1997b), S. 1-2, 5.

<sup>2</sup> Vgl. ASARE/RIGHT (1997b), S. 2.

<sup>3</sup> Vgl. ASARE/RIGHT (1997b), S. 5.

<sup>4</sup> Vgl. ASARE/RIGHT (1997a), S. 740-741.

<sup>5</sup> Weitergehende Erläuterungen zur beschränkten Rationalität finden sich in Kap. 3.2.2.2.2.1.

<sup>6</sup> Vgl. ASARE/RIGHT (1997a), S. 740.

hypothese vorzunehmen.<sup>1</sup> HEIMAN stellte fest, dass die Generierung zusätzlicher Hypothesen einen signifikanten Einfluss auf die Bewertung der Zielhypothese ausübt. Die Prüfer reduzierten ihre Wahrscheinlichkeitsbeurteilung der Zielhypothese um durchschnittlich 9,9 Prozentpunkte.<sup>2</sup> Demnach weisen die Befunde darauf hin, dass Prüfer unter Berücksichtigung zusätzlicher Abweichungsursachen ihre ursprüngliche Hypothesenbewertung anpassen, was die Verwendung eines komplementären Bewertungsverfahrens vermuten lässt.

Allerdings ist der Studie von HEIMAN entgegen zu halten, dass sie sich lediglich auf die Wahrscheinlichkeitsanpassung der Zielhypothese konzentriert hat. Da sie jedoch nicht die Beurteilung der übrigen Hypothesen in die Untersuchung einbezogen hat, konnte letztendlich kein direkter Beweis für die Verwendung eines unabhängigen Verfahrens erbracht werden. Um eine methodische Korrektur bemüht, konzipierten ASARE/WRIGHT eine vergleichbare Studie, welche die Bewertung aller zur Verfügung stehenden Hypothesen explizit berücksichtigt. Die Autoren ließen 62 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 4,3 Jahren eine unerwartete Abweichung in der Bruttogewinnmarge untersuchen.<sup>3</sup> Den Probanden wurden dabei fünf potenzielle Abweichungsursachen mit der Bitte zur Verfügung gestellt, für alle Hypothesen eine Wahrscheinlichkeitsbeurteilung vorzunehmen. Anschließend erhielten die Prüfer weitere auf die generierten Hypothesen ausgerichtete Prüfungsnachweise. Nachdem sie die neuen Erkenntnisse erlangt hatten, war es den Prüfern möglich, ihre ursprüngliche Hypothesenbewertung anzupassen.<sup>4</sup>

Im Ergebnis konnten die Autoren nachweisen, dass die Modifikation der Zielhypothese überwiegend keine korrespondierende Anpassung der alternativen Hypothesen auslöste. Die Summe aus der Änderung der Eintrittswahrscheinlichkeit der Zielhypothese und der Änderungen der Wahrscheinlichkeiten der alternativen Hypothesen ist signifikant größer als null. Demnach sprechen die Ergebnisse der Studie eindeutig für die Verwendung eines unabhängigen Verfahrens zur Beurteilung von multiplen Hypothesen.<sup>5, 6</sup>

---

<sup>1</sup> Vgl. HEIMAN (1990), S. 880-881.

<sup>2</sup> Vgl. HEIMAN (1990), S. 883.

<sup>3</sup> Vgl. ASARE/WRIGHT (1997b), S. 742.

<sup>4</sup> Vgl. ASARE/WRIGHT (1997b), S. 744-746.

<sup>5</sup> Vgl. ASARE/WRIGHT (1997b), S. 749.

<sup>6</sup> ASARE/WRIGHT (1997a) konnten die Verwendung des unabhängigen Verfahrens in einer weiteren Studie erneut bestätigen. Im Gegensatz zur vorherigen Studie war es den 118 Prüfern mit einer durchschnittlichen Prüfungserfahrung von 3,7 Jahren hingegen möglich, eine interne Hypothesengenerierung sowie eine eigenständige Informationssuche vorzunehmen. Der sonstige Untersuchungsaufbau war ansonsten identisch zur vorherigen Studie.

Für die Prüfungspraxis kann an dieser Stelle keine allgemeingültige Empfehlung ausgesprochen werden. Vielmehr obliegt es ihrer eigenen Beurteilung, welche Vorgehensweise in welchen Situationen zu präferieren ist. Aus Kostengründen wäre eine komplementäre Hypothesenbewertung sicherlich zu bevorzugen. Gesellschaften, die dem Wirtschaftlichkeitsziel eine höhere Bedeutung zuordnen, könnten ihre Mitarbeiter für die Dualität von Prüfungsnachweisen, etwa im Rahmen von Schulungen, sensibilisieren. Steht hingegen eine möglichst umfassende Absicherung gegen potenzielle Schadensersatz- und Haftungsansprüche im Vordergrund, empfiehlt es sich, an der Erlangung von direkten Nachweisen anhand der unabhängigen Hypothesenbewertung festzuhalten. Möglich wäre allerdings auch, die Wahl des zu verwendenden Verfahrens an die im Zusammenhang mit dem jeweiligen Auftrag stehende notwendige Prüfungssicherheit zu knüpfen. So wäre etwa vorstellbar, dass bei der Durchführung von analytischen Prüfungshandlungen im Rahmen von prüferischen Durchsichten, bei denen in der Regel eine geringere Prüfungssicherheit notwendig ist, das komplementäre Verfahren zu verwenden wäre, bei „gewöhnlichen“ Jahresabschlussprüfung hingegen das unabhängige Verfahren.<sup>1</sup>

#### **3.5.5.4.3 Die Konjunktionsregel bei der Beurteilung mehrerer Hypothesen**

Möchte der Prüfer eine Mehrzahl von potenziellen Abweichungsursachen berücksichtigen, um zu einem Wahrscheinlichkeitsurteil zu gelangen, so muss er aus statistischer Sicht dabei die Konjunktionsregel befolgen. Diese besagt im Allgemeinen, dass der gemeinsame Eintritt zweier Ereignisse nicht wahrscheinlicher sein kann als die Wahrscheinlichkeit eines der beiden Ereignisse.<sup>2</sup>

In der empirischen Prüfungsforschung waren es erstmalig HO/MAY, die der Frage nachgingen, ob Prüfer bei der Durchführung von analytischen Prüfungshandlungen die Konjunktionsregel einhalten bzw. verletzen. Hierzu ließen sie insgesamt 116 Prüfer mit einer Prüfungserfahrung von durchschnittlich 3,2 Jahren eine Abweichung in der Bruttogewinnmarge untersuchen.<sup>3</sup> Um herauszufinden, welche Bedingungen die Nicht-Befolgung der Konjunktionsregel begünstigen, variierten die Autoren in ihrer Untersuchung (1) die Extremität der zu beurteilenden Abweichung, (2) die Plausibilität der bei-

---

<sup>1</sup> Eine ausführliche Definition der prüferischen Durchsicht findet sich bei MARTEN/QUICK/RUHNKE (2006), S. 580.

<sup>2</sup> Eine detaillierte Beschreibung der Konjunktionsregel erfolgt in Kap. 3.2.2.2.2.2.

<sup>3</sup> Vgl. HO/MAY (1993), S. 86-87.



den in Betracht zu ziehenden Abweichungsgründe sowie (3) die Art der Abweichungsursache (Fehler- oder Nicht-Fehlerursache).<sup>1</sup> Dabei stellte sich heraus, dass die Extremität der Abweichung<sup>2</sup> einerseits und das Zusammenspiel zwischen der Plausibilität<sup>3</sup> und der Art der zu berücksichtigenden Abweichungsursachen andererseits einen signifikanten Einfluss auf die Häufigkeit ausübte, mit der die Konjunktionsregel verletzt bzw. eingehalten wurde.

Hatten die Prüfer eine extreme Abweichung zu beurteilen, so verletzen sie die Konjunktionsregel signifikant häufiger als bei der Beurteilung einer nicht extremen Abweichung (75% im Vgl. zu 43%).<sup>4</sup> Die Autoren führten dieses darauf zurück, dass Prüfer die Verursachung einer extremen Abweichung durch lediglich einen Grund nicht als ausreichend erachteten bzw. zwei Abweichungsgründe als repräsentativer empfanden. Beim Vorliegen einer nicht-extremen Abweichung empfanden sie die Begründung durch eine einzige Ursache hingegen als ausreichend.<sup>5</sup>

Handelte es sich bei den beiden zu beurteilenden Abweichungsursachen um eine hoch plausible Fehlerursache einerseits und eine hoch plausible Nicht-Fehlerursache andererseits, so verletzten die Probanden ebenfalls signifikant häufiger die Konjunktionsregel (76%) als wenn es sich um zwei hoch plausible Fehlerursachen (50%), um zwei weniger plausible Fehlerursachen (51%) bzw. um eine weniger plausible Fehlerursache zum einen und eine weniger plausible Nicht-Fehlerursache zum anderen (59%) handelte.<sup>6</sup> Die Autoren vermuteten, dass Prüfer in der Prüfungspraxis wohl äußerst selten auf das gleichzeitige Vorliegen zweier hoch plausibler Fehlerursachen stoßen und daher die Kombination aus einer hoch plausiblen Fehlererklärung und einer hoch plausiblen Nicht-Fehlererklärung als repräsentativer erachten.<sup>7</sup> Hierfür spricht auch die an anderer Stelle bereits erwähnte Vermutung, dass Prüfer in der Praxis nur selten mit wesentlichen Fehlern konfrontiert werden.<sup>8</sup>

---

<sup>1</sup> Vgl. HO/MAY (1993), S. 81-84.

<sup>2</sup> In einer extremen Fallvariante betrug die Abweichung zwischen Soll- und Ist-Wert 6,8 Prozentpunkte im Vergleich zu 1,3 Prozentpunkten in der nicht-extremen Fallkonstellation. Vgl. HO/MAY (1993), S. 88.

<sup>3</sup> Sowohl hoch plausible als auch weniger plausible Abweichungsursachen kommen theoretisch in Betracht. Allerdings sind Abweichungsursachen mit einer hohen Plausibilität im Gegensatz zu weniger plausiblen Ursachen durch eine höhere Beobachtungshäufigkeit in der Prüfungsrealität gekennzeichnet. Vgl. hierzu HO/MAY (1993), S. 88-89.

<sup>4</sup> Vgl. HO/MAY (1993), S. 91.

<sup>5</sup> Vgl. HO/MAY (1993), S. 92.

<sup>6</sup> Vgl. HO/MAY (1993), S. 93.

<sup>7</sup> Vgl. HO/MAY (1993), S. 93.

<sup>8</sup> Siehe hierzu etwa Kap. 3.5.5.2.

Sollten in der Praxis tätige Prüfer ebenfalls unter gewissen Bedingungen tendenziell dazu neigen, dem Konjunktionsfehler zu unterliegen, so könnte diese Verzerrung die Effizienz der analytischen Prüfungshandlung und somit der gesamten Jahresabschlussprüfung erheblich beeinträchtigen, etwa wenn weitere Untersuchungen auf zwei potenzielle Fehlerbereiche vorgenommen werden, tatsächlich jedoch nur eine Fehlerursache zu der festgestellten Abweichung geführt hat. Daher empfiehlt es sich, Prüfer im Rahmen von Fortbildungsmaßnahmen auf die mögliche Gefährdung durch einen Konjunktionsfehler aufmerksam zu machen und sie aufzufordern, sich die Konjunktionsregel im Rahmen ihrer prüferischen Tätigkeiten regelmäßig zu vergegenwärtigen.

### 3.5.5.5 Tabellarische Darstellung der empirischen Befunde zur Phase der Hypothesenbewertung

| Autor(en)                      | Teilnehmer |                     | Daten-<br>gewinnung | Daten-<br>auswertung | Ergebnisse                                                                                                                                                                                                                                                                                                                        | Signifikanz |
|--------------------------------|------------|---------------------|---------------------|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
|                                | Anzahl     | Art                 |                     |                      |                                                                                                                                                                                                                                                                                                                                   |             |
| KAPLAN/RECKERS<br>(1989)       | 71         | k.A.<br>Ø = 8 J.    | Experiment          | Varianzanalyse       | Die Anzahl der Berufsjahre hat einen signifikanten Einfluss auf die Erwartungen darüber, ob der Grund für eine festgestellte Abweichung auf eine Fehler- oder Nicht-Fehlerursache zurückzuführen ist. Unerfahrene Prüfer gehen verstärkt von einer Fehlerursache aus; erfahrenere Prüfer verstärkt von einer Nicht-Fehlerursache. | 0,005       |
| GREEN/TROTMAN<br>(2003)        | 82         | k.A.<br>Ø = 3,8 J.  | Experiment          | Häufigkeiten         | 34% der Prüfer mit Mandantenerklärung ordneten der erhaltenen Mandantenerklärung die höchste Eintrittswahrscheinlichkeit zu, obwohl bereits Informationen vorlagen, die auf die Inkorrektheit der Erklärung hindeuteten.                                                                                                          | k.A.        |
|                                |            |                     |                     | Varianzanalyse       | Dabei fielen ihre Wahrscheinlichkeitseinschätzungen signifikant höher aus als die der anderen Prüfer.                                                                                                                                                                                                                             | 0,000       |
| GREEN (2004b)                  | 58         | k.A.<br>Ø = 4 J.    | Experiment          | Häufigkeiten         | 42% der Prüfer mit Mandantenerklärung ordneten der Mandantenerklärung die höchste Eintrittswahrscheinlichkeit zu, obwohl bereits Informationen vorlagen, die auf die Inkorrektheit der Erklärung hindeuteten. Demgegenüber ordnete keiner der Prüfer ohne Mandantenerklärung dieser die höchste Eintrittswahrscheinlichkeit zu.   | k.A.        |
|                                |            |                     |                     | Chi-Quadrat-Test     | Prüfer mit Mandantenerklärung wählten diese signifikant häufiger als Fehlerursache aus als Prüfer ohne Mandantenerklärung.                                                                                                                                                                                                        | 0,002       |
| ASARE/WRIGHT/<br>WRIGHT (1998) | 73         | k.A.<br>Ø = 4,95 J. | Experiment          | Häufigkeiten         | 67,5% der Prüfer mit Mandantenerklärung ordneten der Mandantenerklärung die höchste Eintrittswahrscheinlichkeit zu, obwohl bereits Informationen vorlagen, die auf die Inkorrektheit der Erklärung hindeute-                                                                                                                      | k.A.        |

|                        |    |                              |            |                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                  |
|------------------------|----|------------------------------|------------|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
|                        |    |                              |            | Wilcoxon-Vorzeichen-Rang-Test    | ten. Demgegenüber ordneten 43,7% der Prüfer ohne Mandantenerklärung dieser die höchste Eintrittswahrscheinlichkeit zu.<br>Prüfer mit Mandantenerklärung wählten diese signifikant häufiger als Fehlerursache aus als Prüfer ohne Mandantenerklärung.                                                                                                                                                                                                                                                                                                                                                           | 0,0217           |
| ANDERSON/KOONCE (1995) | 60 | k.A.<br>Ø = 2 J.             | Experiment | Häufigkeiten<br><br>Häufigkeiten | Prüfer mit einer teilweise inkorrekten Mandantenerklärung schätzten die Wahrscheinlichkeit der Richtigkeit und Vollständigkeit der Mandantenerklärung fast identisch ein wie Prüfer mit korrekter Mandantenerklärung. Der überwiegende Teil der Prüfer, die der inkorrekten Mandantenerklärung die höchste Eintrittswahrscheinlichkeit zuordneten, überprüften lediglich die Plausibilität dieser Hypothese. Demgegenüber führte der Großteil der Prüfer, die die Inkorrektheit bzw. Unvollständigkeit der Mandantenerklärung erkannten zusätzlich eine rechnerische Überprüfung durch (sog. Quantifizierung). | k.A.<br><br>k.A. |
| KOONCE (1992)          | 26 | k.A.<br>Ø = 2,8 J.           | Experiment | t-Test                           | Prüfer, die Argumente für die Gültigkeit der Mandantenerklärung schriftlich dokumentierten, ordneten dieser eine signifikant höhere Eintrittswahrscheinlichkeit zu als Prüfer, die keine Argumente dokumentierten (sog. written explanation effect).                                                                                                                                                                                                                                                                                                                                                           | 0,0112           |
| KOONCE/PHILLIPS (1996) | 38 | Prüfungsleiter<br>Ø = 2,9 J. | Experiment | Mittelwerte                      | Prüfer schätzten die Wahrscheinlichkeit der Korrektheit einer inkorrekten Mandantenerklärung zunächst durchschnittlich auf 44,1% ein. Nachdem sie Argumente für die Korrektheit der Mandantenerklärung schriftlich niederlegten, korrigierten sie ihre Wahrscheinlichkeitseinschätzung auf durchschnittlich 78,7%.                                                                                                                                                                                                                                                                                             | k.A.             |
| KOONCE (1992)          | 26 | k.A.<br>Ø = 2,8 J.           | Experiment | t-Test                           | Prüfer, die neben Argumenten für die Gültigkeit der Mandantenerklärung auch Gegenargumente dokumentierten, schätzten die Wahrscheinlichkeit der Man-                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0,0015           |

|                                          |     |                               |            |                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                    |
|------------------------------------------|-----|-------------------------------|------------|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
|                                          |     |                               |            |                      | dantenerklärung signifikant geringer ein als Prüfer, die ausschließlich Argumente für die Gültigkeit der Mandantenerklärung aufführten (sog. counterexplanation effect).                                                                                                                                                                                                                                                                                    |                    |
| ANDER-<br>SON/KOONCE<br>(1998)           | 36  | k.A.<br>Ø = 42 Mon.           | Experiment | Chi-Quadrat-Test     | Prüfer mit Quantifizierung erkannten die Inkorrektheit der Mandantenerklärung signifikant häufiger als Prüfer, die keine Quantifizierung durchführten.                                                                                                                                                                                                                                                                                                      | 0,0425             |
| ANDER-<br>SON/KADOUS/<br>KOONCE (2004)   | 113 | Prüfungsleiter<br>Ø = 41 Mon. | Experiment | Varianzanalyse       | Prüfer, die neben der Mandantenerklärung auch eine Quantifizierung vom Mandanten erhielten, schätzten die Eintrittswahrscheinlichkeit der Mandantenerklärung nicht signifikant höher ein, als Prüfer, die keine Quantifizierung vom Mandanten erhielten.                                                                                                                                                                                                    | 0,24               |
| KOONCE/PHILLIPS<br>(1996)                | 38  | Prüfungsleiter<br>Ø = 2,9 J.  | Experiment | t-Test<br><br>t-Test | Prüfer mit einer einfach nachzuvollziehenden Information schätzen die Eintrittswahrscheinlichkeit der Mandantenerklärung signifikant höher ein als die Prüfer mit den schwierig nachzuvollziehenden Informationen. Prüfer mit einer einfach nachzuvollziehenden Information schätzen die Eintrittswahrscheinlichkeit einer anderen Abweichungshypothese nicht signifikant niedriger ein als die Prüfer mit den schwierig nachzuvollziehenden Informationen. | 0,0001<br><br>0,15 |
| ANDER-<br>SON/KOONCE/<br>MARCHANT (1994) | 124 | k.A.<br>Ø = 74 Mon.           | Experiment | Varianzanalyse       | Die Kompetenz der Informationsquelle (hier Controller) hat einen signifikanten Einfluss auf die Wahrscheinlichkeits einschätzung der Mandantenerklärung. Prüfer, die eine Mandantenerklärung von einem kompetenten Controller erhielten, schätzten deren Eintrittswahrscheinlichkeit signifikant höher ein als Prüfer, die eine Mandantenerklärung von einem weniger kompetenten Controller erhielten.                                                      | 0,0335             |
| PEECHER (1996)                           | 106 | k.A.<br>Ø = 3 J.              | Experiment | Varianzanalyse       | Die Integrität der Informationsquelle hat einen signifikanten Einfluss auf die Wahrscheinlichkeits einschätzung der Mandantenerklärung. Prüfer, die eine Erklärung von einem Mandanten mit hoher Integrität erhielt                                                                                                                                                                                                                                         | 0,00               |

|                                |     |                                                                        |            |                  |                                                                                                                                                                                                                                                                                                                                                                  |       |
|--------------------------------|-----|------------------------------------------------------------------------|------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                                |     |                                                                        |            |                  | ten, schätzten deren Eintrittswahrscheinlichkeit signifikant höher ein als Prüfer, die eine Erklärung von einem Mandanten mit geringerer Integrität erhielten.                                                                                                                                                                                                   |       |
| GREEN (2004a)                  | 63  | k.A.<br>Ø = 3,84 J.                                                    | Experiment | Häufigkeiten     | Prüfer, denen eine inkorrekte Mandantenerklärung mitgeteilt wurde, weisen eine vergleichbare Erfolgsquote auf wie Prüfer, denen keine Mandantenerklärung mitgeteilt wurde.                                                                                                                                                                                       | k.A.  |
| ANDERSON/KAPLAN/RECKERS (1997) | 184 | 104 Prüfungsleiter<br>Ø = 38,4 Mon.<br><br>80 Manager<br>Ø = 60,7 Mon. | Experiment | Varianzanalyse   | Die Bereitstellung einer Entscheidungshilfe, die überwiegend Fehlerhypothesen enthält (sog. Fehler-Dominanz), hat einen signifikanten Einfluss auf die Beurteilung der Fehlerwahrscheinlichkeit. Prüfer, denen eine fehlerdominierte Liste zur Verfügung gestellt wurde, nahmen eine signifikante Erhöhung der Fehlerwahrscheinlichkeit vor.                     | 0,001 |
|                                |     |                                                                        |            | Varianzanalyse   | Die Bereitstellung einer Entscheidungshilfe, die überwiegend Nicht-Fehlerhypothesen enthält (sog. Nicht-Fehler-Dominanz), hat keinen signifikanten Einfluss auf die Beurteilung der Fehlerwahrscheinlichkeit. Prüfer, denen eine nicht-fehlerdominierte Liste zur Verfügung gestellt wurde, nahmen keine signifikante Erhöhung der Fehlerwahrscheinlichkeit vor. | 0,44  |
| ANDERSON/MORENO/MUELLER (2003) | 51  | K.A.<br>Ø = 74 Mon.                                                    | Experiment | Kovarianzanalyse | Prüfer, denen eine Abweichungshypothese von einer Entscheidungshilfe zur Verfügung gestellt wird, schätzen ihre Eintrittswahrscheinlichkeit signifikant höher als Prüfer, die eine identische Abweichungshypothese                                                                                                                                               | 0,042 |

|                      |     |                    |            |                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                        |
|----------------------|-----|--------------------|------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
|                      |     |                    |            |                     | von dem Mandanten erhalten.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                        |
| HEIMAN (1990)        | 73  | k.A.<br>Ø = 3,5 J. | Experiment | Varianzanalyse      | Die Generierung zusätzlicher Hypothesen hat einen signifikanten Einfluss auf die Wahrscheinlichkeitsbeurteilung der Zielhypothese (sog. komplementäres Verfahren). Prüfer, die zusätzliche Hypothesen zur Verfügung gestellt bekommen bzw. selbst generieren, bewerten die Eintrittswahrscheinlichkeit der Zielhypothese anschließend signifikant geringer ein.                                                                                                                                                                                                                                                                                                                                          | $< 0,01^1 / = 0,058^2$ |
| ASARE/WRIGHT (1997b) | 62  | k.A.<br>Ø = 4,3 J. | Experiment | Varianzanalyse      | Die Modifikation der Zielhypothese bringt keine Anpassung der Eintrittswahrscheinlichkeit der alternativen Hypothesen mit sich (sog. unabhängiges Verfahren). Die Summe aus der Änderung der Eintrittswahrscheinlichkeit der Zielhypothese und der Änderungen der Wahrscheinlichkeiten der alternativen Hypothesen ist signifikant größer als null.                                                                                                                                                                                                                                                                                                                                                      | $< 0,05$               |
| HO/MAY (1993)        | 116 | k.A.<br>3,2 J.     | Experiment | Log-Lineare Analyse | Die Extremität einer Abweichung einerseits und das Zusammenspiel zwischen Plausibilität und Art der Abweichungsursachen andererseits haben einen signifikanten Einfluss auf die Häufigkeit, mit der die Konjunktionsregel verletzt bzw. eingehalten wird. Prüfer, die eine extreme Abweichung zu berücksichtigen hatten, verletzten die Konjunktionsregel häufiger als Prüfer, die keine extreme Abweichung zu berücksichtigen hatten. Prüfer, die eine hoch plausible Fehlerursache und eine hoch plausible Nicht-Fehlerursache zu berücksichtigen hatten, verletzten die Konjunktionsregel signifikant häufiger als Prüfer, die zwei hoch plausible Fehlerursachen, Prüfer, die zwei weniger plausible | $0,001^1 / 0,044^2$    |

<sup>1</sup> Signifikanzniveau bei Prüfern, denen anschließend zusätzliche Hypothesen mitgeteilt wurden.

<sup>2</sup> Signifikanzniveau bei Prüfern, die anschließend mindestens zwei weitere Hypothesen selbst generierten.

|  |  |  |  |  |                                                                                                                                               |  |
|--|--|--|--|--|-----------------------------------------------------------------------------------------------------------------------------------------------|--|
|  |  |  |  |  | Fehlerursachen und Prüfer, die eine weniger plausible Fehlerursache und eine weniger plausible Nicht-Fehlerursache zu berücksichtigen hatten. |  |
|--|--|--|--|--|-----------------------------------------------------------------------------------------------------------------------------------------------|--|

Tabelle 5: Ergebnisse der empirischen Prüfungsforschung zur Phase der Hypothesenbewertung

---

<sup>1</sup> Signifikanzniveau hinsichtlich der Extremität der Abweichung.

<sup>2</sup> Signifikanzniveau hinsichtlich des Zusammenspiels zwischen der Plausibilität und der Art der Abweichungsursachen.



### **3.5.6 Studien zur Phase der Prüfungsentscheidung**

#### **3.5.6.1 Vorbemerkung**

In dieser abschließenden Phase einer analytischen Prüfungshandlung obliegt es dem Prüfer, eine Entscheidung hinsichtlich der weiteren prüferischen Vorgehensweise zu treffen. Von immenser Bedeutung ist an dieser Stelle die Frage, in welcher Phase die analytische Prüfungshandlung eingesetzt wurde.<sup>1</sup> Wurde von ihr im Rahmen der Prüfungsdurchführung Gebrauch gemacht, so wird die Prüfungsentscheidung im Wesentlichen durch das Resultat der Hypothesenbewertung determiniert. Geht der Prüfer davon aus, dass es sich bei der Abweichungsursache um einen Fehlergrund handelt, so wird er den Mandanten zur Vornahme einer Nachbuchung veranlassen. Ist der Prüfer hingegen der Ansicht, dass die Abweichung durch einen Nicht-Fehler verursacht wurde, so kann er das Ergebnis der analytischen Prüfungshandlung inkl. der zur Identifikation der Nicht-Fehlerursache vorgenommenen Prüfungshandlungen als Prüfungsnachweis für die Korrektheit des geprüften Prüffelds verwenden.

Zu unterscheiden ist hiervon die Anwendung einer analytischen Prüfungshandlung im Rahmen der Prüfungsplanung. Zu diesem noch sehr frühen Zeitpunkt der Jahresabschlussprüfung kann das Ziel einer analytischen Prüfungshandlung nicht in der Identifikation des Abweichungsgrunds liegen. Vielmehr nimmt der Prüfer das Vorliegen bzw. Nicht-Vorliegen einer wesentlichen Abweichung zum Anlass, seine personellen und sachlichen Ressourcen möglichst zweckdienlich zu verteilen.<sup>2</sup> Im Mittelpunkt steht hierbei die Erstellung des Prüfungsplans. Analytische Prüfungshandlungen helfen ihm hierbei, das Fehlerrisiko einzelner Prüfungsbereiche zu beurteilen und Prüfungsschwerpunkte zu identifizieren.

Da der Handlungsspielraum beim Treffen einer Prüfungsentscheidung im Rahmen der Phase der Prüfungsdurchführung äußerst eingeschränkt ist (etwas vereinfacht zum Ausdruck gebracht, kann sich der Prüfer hier lediglich zwischen einer bzw. keiner Nachbuchung entscheiden), verwundert es nicht, dass sich die empirischen Forschungsaktivitäten ausschließlich der Anwendung von analytischen Prüfungshandlungen im Rahmen

---

<sup>1</sup> Siehe zu den unterschiedlichen Einsatzbereichen von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung ausführlich Kap. 2.1.2.

<sup>2</sup> Ähnlich argumentieren HIRST/KOONCE (1996), S. 468.

der Prüfungsplanung gewidmet haben. Dabei kann zwischen experimentellen Studien einerseits und fragebogengestützten Studien andererseits differenziert werden.<sup>1</sup>

### **3.5.6.2 Die Prüfungsentscheidung im Rahmen des Einsatzes zur Prüfungsplanung**

#### **3.5.6.2.1 Experimentelle Studien**

BIGGS/MOCK/WATKINS untersuchten in ihrer explorativen Studie zur Durchführung von analytischen Prüfungshandlungen u.a. auch die Prüfungsentscheidung der teilnehmenden Probanden, nachdem verschiedene analytische Prüfungshandlungen durchgeführt wurden.<sup>2</sup> Die Autoren forderten die Prüfer auf, auf Basis der jeweiligen Ergebnisse eine Modifikation des Prüfungsplans vorzunehmen, wobei die analytischen Prüfungshandlungen im Bereich der Umsätze & Forderungen sowohl Abweichungen als auch Nicht-Abweichungen umfassten. Die Autoren stellten dabei fest, dass von den insgesamt 422 Modifikationen lediglich 19 eine Reduzierung und 403 eine Erhöhung des Prüfungsumfangs darstellten.<sup>3</sup> BIGGS/MOCK/WATKINS gingen davon aus, dass die Prüfer die Ergebnisse der analytischen Prüfungshandlungen lediglich als Warnhinweise für besondere Risiken (sog. red flags) nutzen. Die Möglichkeit, den Prüfungsumfang einzuschränken, wenn keine wesentliche Abweichung identifiziert werden konnte, wurde von den Prüfern hingegen nicht wahrgenommen. Dieses Phänomen wird in der Prüfungsforschung auch als *asymmetrischer Einsatz* bezeichnet.<sup>4, 5</sup>

COHEN/KIDA führten diesen asymmetrischen Einsatz auf die berufs- bzw. ausbildungsbedingte professionelle Skepsis von Abschlussprüfern zurück. Da BIGGS/MOCK/WATKINS in ihrer Studie jedoch lediglich die Vorgehensweise von fünf Prüfern beobachteten, untersuchten COHEN/KIDA, ob ein asymmetrischer Einsatz auch bei einer größeren Anzahl von Prüfern festgestellt werden kann. Daher ließen sie 96 Prüfer eine vergleichbare Aufgabe bearbeiten wie BIGGS/MOCK/WATKINS. Um darüber hinaus feststellen zu können, ob der asymmetrische Einsatz möglicherweise durch be-

---

<sup>1</sup> Zur Darstellung der unterschiedlichen Forschungsmethoden sei auf Kap. 3.4 verwiesen.

<sup>2</sup> Eine ausführliche Beschreibung des Untersuchungsaufbaus findet sich unter Kap. 3.5.2.1.

<sup>3</sup> Vgl. BIGGS/MOCK/WATKINS (1988), S. 156.

<sup>4</sup> Vgl. BIGGS/MOCK/QUICK (2000), S. 173.

<sup>5</sup> Dieses Phänomen wurde auch im Rahmen einer empirisch-archivistischen Studie von BEDARD nachgewiesen. Sie stellte bei der Analyse der Arbeitspapiere von 48 Prüfungsaufträgen fest, dass es sich bei allen Prüfungsplanänderungen, die auf Basis der Ergebnisse von analytischen Prüfungshandlungen veranlasst wurden, um Erhöhungen des Prüfungsumfangs handelte. Vgl. BEDARD (1989), S. 59, 66.

stimmte externe Faktoren beeinflusst wird, ließen die Autoren einen Teil der Probanden Hinweise zukommen, die auf ein erhöhtes Kontrollrisiko hindeuteten. Die anderen Prüfer wurden hingegen darauf hingewiesen, dass von einem geringen Kontrollrisiko ausgegangen werden kann.<sup>1</sup>

COHEN/KIDA stellten fest, dass Prüfer den Prüfungsumfang bei der Identifikation einer wesentlichen Abweichung signifikant erhöhten. Somit nutzen die Prüfer auch in dieser Studie die Ergebnisse der analytischen Prüfungshandlungen als red flags, die auf Prüfungsbereiche mit erhöhtem Fehlerrisiko hindeuteten.<sup>2</sup> Darüber hinaus wiesen die Ergebnisse ebenfalls darauf hin, dass Prüfer analytische Prüfungshandlungen nicht zur Reduzierung des Prüfungsumfangs einsetzten. Prüfer, die keine wesentliche Abweichung identifizieren konnten, nahmen keine Reduzierung des Prüfungsumfangs vor. Vielmehr wurden die Prüfungshandlungen um durchschnittlich 18,9% erhöht. Dieses konservative Verhalten konnte sogar beobachtet werden, wenn ein geringes Kontrollrisiko vorlag. Im Schnitt stieg der Prüfungsumfang hier um 10,1%.<sup>3, 4</sup> Insofern konnten der asymmetrische Einsatz auch durch COHEN/KIDA bestätigt werden.

Da sich sowohl die Studie von BIGGS/MOCK/WATKINS als auch von COHEN/KIDA auf den Bereich der Umsätze & Forderungen bezog, untersuchte COHEN in einer weiteren Studie, ob sich der asymmetrische Einsatz auch im Bereich der Vorräte & Umsatzkosten nachweisen lässt.<sup>5</sup> Im Gegensatz zu den beiden vorherigen Studien variierte COHEN bei der wesentlichen Abweichung zwischen einer Über- und Unterbewertung der Vorratsbestände. Erstgenanntes führte zu einem zu gering ausgewiesenen Jahresüberschuss, Letztgenanntes hingegen zu einem erhöhten Jahresüberschuss.<sup>6</sup>

Insgesamt konnte COHEN den asymmetrischen Einsatz von analytischen Prüfungshandlungen auch für den Bereich der Vorräte & Umsatzkosten bestätigen. Prüfer, die eine wesentliche Abweichung identifizierten, nahmen eine signifikante Erhöhung des Prüfungsumfangs vor und dieses unabhängig davon, ob es sich bei der Abweichung um ei-

---

<sup>1</sup> Vgl. COHEN/KIDA (1989), S. 265-266.

<sup>2</sup> Vgl. COHEN/KIDA (1989), S. 270-271.

<sup>3</sup> Vgl. COHEN/KIDA (1989), S. 271.

<sup>4</sup> Dass Prüfer den Prüfungsumfang trotz fehlender Abweichungen und niedrigem Kontrollrisiko nicht reduzieren, steht hingegen im Widerspruch zu den Ergebnissen von HIRST/KOONCE. Sie interviewten insgesamt 16 Prüfer hinsichtlich ihrer Anwendung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung. Dabei gaben neun Prüfer an, dass sie unter der Voraussetzung, dass die analytische Prüfungshandlung keine wesentliche Abweichung indiziert und zudem kein erhöhtes Kontrollrisiko vorliegt, den Prüfungsumfang reduzieren. Vgl. HIRST/KOONCE (1996), S. 464.

<sup>5</sup> Die Umsätze & Forderungen stellen ebenso wie die Vorräte & Umsatzkosten ein zentrales Prüffeld im Rahmen der Jahresabschlussprüfung dar. Vgl. COHEN (1994), S. 177.

<sup>6</sup> Vgl. COHEN (1994), S. 171-173.

ne Unter- oder Überbewertung handelte (18,3%-ige bzw. 19,6%-ige Erhöhung des Prüfungsumfangs).<sup>1</sup> Hingegen konnte bei den Prüfern, denen keine wesentliche Abweichung zur Verfügung gestellt wurde, keine signifikante Reduzierung des Prüfungsumfangs identifiziert werden.<sup>2</sup> Insofern konnte auch hier der asymmetrische Einsatz erneut bestätigt werden.<sup>3</sup>

Zusammenfassend kann also konstatiert werden, dass alle experimentellen Studien auf einen asymmetrischen Einsatz von analytischen Prüfungshandlungen hindeuten.<sup>4</sup>

### 3.5.6.2.2 Fragebogengestützte Studien

Neben den zuvor dargestellten experimentellen Studien, bei denen die teilnehmenden Probanden einen (möglichst) realitätsnahen Fall zu bearbeiten hatten, bemühte sich ein anderer Teil der empirischen Prüfungsforschung, die Anwendung von analytischen Prüfungshandlungen in der Prüfungspraxis anhand von fragebogengestützten Studien zu untersuchen. FRASER/HATHERLY/LIN etwa analysierten 366 von Prüfern bearbeitete Fragebögen und erkannten entgegen der Erkenntnisse aus den experimentellen Studien, dass mehr als die Hälfte der Prüfer generell dazu bereit ist, den Prüfungsumfang zu reduzieren, falls das Ergebnis der analytischen Prüfungshandlung keine wesentliche Abweichung anzeigt.<sup>5</sup> Diese generelle Bereitschaft zu einem *symmetrischen Einsatz* von analytischen Prüfungshandlungen äußerten auch die 307 befragten Prüfer in der Studie von MULLIGAN/INKSTER.<sup>6</sup> Die Autoren stellten fest, dass 29% der Befragten den Prüfungsumfang immer und 35% der Befragten häufig reduzieren, falls keine signifikante Abweichung zu erkennen ist.<sup>7</sup> Allerdings gaben die Prüfer auch an, dass sie die Ergeb-

---

<sup>1</sup> Vgl. COHEN (1994), S. 176.

<sup>2</sup> Vgl. COHEN (1994), S. 177.

<sup>3</sup> HIRST/KOONCE betrachten die Ergebnisse der experimentellen Studien indes kritisch und weisen darauf hin, dass die Modifikationen der Prüfungsplan in diesen Studien ausschließlich über eine Ausweitung bzw. Reduzierung der Prüfungsstunden gemessen wurde. Im Rahmen ihrer interviewgestützten Studie gaben die Prüfer an, dass sie Anpassungen u.a. auch bei der Art der weiteren Prüfungshandlungen und beim Prüfungspersonal vornehmen, wenn analytische Prüfungshandlungen keine wesentlichen Abweichungen erkennen lassen. Vgl. HIRST/KOONCE (1996), S. 464.

<sup>4</sup> Allerdings steht der asymmetrische Einsatz im Widerspruch zu den Befunden von GLOVER/PRAWITT/WILKS (2005), die im Rahmen ihrer Studie nachweisen konnten, dass Prüfer den Ergebnissen von analytischen Prüfungshandlungen ein besonders hohes Vertrauen entgegenbringen, wenn sie keine Abweichung signalisieren (sog. *favourable outcome bias*). Da die Autoren jedoch nicht die Prüfungsentscheidung untersuchten, sondern die Verlässlicheitseinschätzung des Ergebnisses einer analytischen Prüfungshandlungen, kann ein symmetrischer Einsatz nur vermutet werden. Ein empirischer Nachweis konnte hingegen nicht erbracht werden.

<sup>5</sup> Vgl. FRASER/HATHERLY/LIN (1997), S. 39, 42.

<sup>6</sup> Vgl. MULLIGAN/INKSTER (1999), S. 111.

<sup>7</sup> Vgl. MULLIGAN/INKSTER (1999), S. 113.

nisse der analytischen Prüfungshandlungen erst durch die Ergebnisse anderer Prüfungshandlungen bestätigen lassen, bevor sie den Prüfungsumfang reduzieren.<sup>1</sup> Vergleichbare Erkenntnisse liefert eine weitere Studie von LI/FRASER, in der Fragebögen von 182 Prüfern ausgewertet wurden.<sup>2</sup> 50,5% der Befragten gaben an, dass sie den Prüfungsumfang nur dann reduzieren würden, wenn sich die Ergebnisse der analytischen Prüfungshandlungen durch weitere Prüfungshandlungen bestätigen ließen. 37,4% der Prüfer würden den Umfang auch ohne weitere Prüfungsnachweise einschränken und weitere 12,1% würden den Prüfungsumfang niemals einschränken, wenn analytische Prüfungshandlungen keine Abweichungen erkennen lassen.<sup>3</sup>

Insgesamt betrachtet lassen die Ergebnisse der fragebogengestützten Studien kein eindeutiges Bild über die Verwendung von analytischen Prüfungshandlungen in der Prüfungspraxis erkennen. Eine Bereitschaft zu einem symmetrischen Einsatz scheint zwar generell vorhanden. Allerdings wird eine Einschränkung des Prüfungsumfangs häufig von dem Vorliegen weiterer Prüfungsnachweise abhängig gemacht. Die zur Erlangung von hinreichend sicheren Prüfungsnachweisen kombinierte Durchführung von analytischen Prüfungshandlungen mit anderen Prüfungshandlungen steht zwar grundsätzlich im Einklang mit IDW PS 312.11. Allerdings ist dem Prüfungsstandard auch zu entnehmen, dass analytische Prüfungshandlungen im Rahmen der Prüfungsplanung zur Einschränkung des Prüfungsumfangs herangezogen werden können.<sup>4</sup> Dieser Schritt ist essentiell für die Effizienz von analytischen Prüfungshandlungen, da der Vorteil der Identifikation eines Prüffelds mit geringem Fehlerrisiko nicht in einen wirtschaftlichen Nutzen umgesetzt würde.

#### **3.5.6.2.3 Der Einfluss des Erhalts einer Mandantenerklärung auf die Prüfungsentscheidung**

Wie in allen anderen Phasen der Durchführung von analytischen Prüfungshandlungen stellt sich auch in der Phase der Prüfungsentscheidung die Frage, inwieweit sich der Erhalt einer Mandantenerklärung auf die prüferische Vorgehensweise auswirkt. Allerdings wurde dieser Bereich der empirischen Prüfungsforschung nicht sehr intensiv untersucht. Der Schwerpunkt der Studien zum Einfluss einer Mandantenerklärung liegt insbesonde-

---

<sup>1</sup> Vgl. MULLIGAN/INKSTER (1999), S. 114.

<sup>2</sup> Vgl. LIN/FRASER (2003), S. 157.

<sup>3</sup> Vgl. LIN/FRAZER (2003), S. 161.

<sup>4</sup> Vgl. IDW PS 312.10 i.V.m. 19.

re in den Phasen der Hypothesengenerierung und -bewertung sowie der Informationssuche.<sup>1</sup>

Eine der wenigen Untersuchungen zum Einfluss einer Mandantenerklärung auf die Prüfungsentscheidung führten GLOVER/JIAMBALVO/KENNEDY durch. Sie ließen insgesamt 68 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 3,1 Jahren eine wesentliche Diskrepanz in der Umschlagshäufigkeit der Vorräte sowie der Bruttogewinnmarge analysieren.<sup>2</sup> Alle Prüfer erhielten dabei die identische Erklärung des Mandanten, dass die Abweichung auf einen Nicht-Fehler zurückzuführen sei.<sup>3</sup> Allerdings erhielt eine Gruppe der Prüfer eine zusätzliche Information, welche die Erklärung des Mandanten weitestgehend bestätigte (high corroboration). Der anderen Prüfergruppe wurde hingegen eine Information zur Verfügung gestellt, welche die Mandantenerklärung lediglich geringfügig bestätigte (low corroboration).<sup>4</sup> Zusätzlich variierten GLOVER/JIAMBALVO/KENNEDY das jeweilige Prüfungsrisiko.<sup>5</sup>

Zwar erkannten die Autoren, dass Prüfer, die ein erhöhtes Prüfungsrisiko zu berücksichtigen hatten und zudem lediglich auf eine geringfügige Bestätigung der Mandantenerklärung (low corroboration) zurückgreifen konnten, signifikant häufiger eine Erhöhung des Prüfungsumfangs vornahmen als die übrigen Prüfer (56% im Vgl. zu 25% bei geringem Risiko und low corroboration, im Vgl. zu 23,5% bei geringem Risiko und high corroboration, im Vgl. zu 12,5% bei erhöhtem Risiko und high corroboration). Dennoch ist bedenklich, dass bald jeder zweite Prüfer (44%) trotz dieser äußerst kritischen Situation (hohes Prüfungsrisiko und low corroboration) nicht dazu bereit war, den Prüfungsumfang zu erhöhen.<sup>6, 7</sup>

---

<sup>1</sup> Vgl. hierzu ausführlich Kap. 3.5.3 für die Hypothesengenerierung, Kap. 3.5.5 für die Hypothesenbewertung sowie Kap. 3.5.4 für die Informationssuche.

<sup>2</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 31.

<sup>3</sup> Die genaue Begründung des Mandanten lag in der Reduzierung des Schwunds in den Vorräten aufgrund der Einführung von effektiven Kontrollen sowie der Einräumung von Preisnachlässen zweier Großlieferanten aufgrund erfolgreicher Vertragsverhandlungen. Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 32, 33.

<sup>4</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 33.

<sup>5</sup> Den Gruppen mit erhöhtem Prüfungsrisiko wurde mitgeteilt, dass der Mandant aufgrund eines vorgesehenen Unternehmensverkaufs den Anreiz hat, einen möglichst hohen Jahresüberschuss auszuweisen. Den anderen Prüfern wurde diese Information hingegen nicht mitgeteilt. Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 32, 33.

<sup>6</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 38-39.

<sup>7</sup> Allerdings ist zu erwähnen, dass der von den Probanden zu modifizierende Prüfungsplan bereits mit dem Wissen des erhöhten Prüfungsrisikos erstellt wurde. Dieses könnte dazu geführt haben, dass ein Großteil der Prüfer bereits einen umfassenden Prüfungsplan erstellt, der unabhängig davon, ob die Mandantenerklärung weitestgehend bestätigt wurde, noch nicht auf die Erreichung einer höchstmöglichen Prüfungssicherheit ausgerichtet ist. Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 35 i.V.m. 37.

Zudem konnten die Autoren feststellen, dass Prüfer mit niedrigem Prüfungsrisiko und geringfügiger Bestätigung der Mandantenerklärung (low corroboration) nicht signifikant häufiger eine Erhöhung des Prüfungsumfangs durchführten als Prüfer, die ein geringes Prüfungsrisiko zu berücksichtigen hatten, zudem jedoch auf eine weitgehende Bestätigung der Mandantenerklärung (high corroboration) zurückgreifen konnten (25% im Vgl. zu 23,5%).<sup>1</sup> Dieser Befund lässt darauf schließen, dass Prüfer bei geringem Prüfungsrisiko einer geringfügig bestätigten Mandantenerklärung ein ähnlich hohes Vertrauen schenken wie einer umfassend bestätigten Mandantenerklärung.

Schließlich war auch das Verhalten derjenigen Prüfer bedenklich, die einem hohen Prüfungsrisiko unterlagen und die Mandantenerklärung weitgehend bestätigt wurde (high corroboration). Hier führten lediglich 12,5% der Prüfer eine Erhöhung des Prüfungsumfangs durch.<sup>2</sup> Damit weicht die Häufigkeit, mit der diese Prüfer den Prüfungsumfang erhöhten, nicht signifikant von der Häufigkeit ab, mit der Prüfer den Umfang erhöhten, wenn ein niedriges Prüfungsrisiko und eine weitgehende Bestätigung der Mandantenerklärung (high corroboration) zu berücksichtigen war. Damit erscheint auch hier das Vertrauen der Prüfer in die Korrektheit der Mandantenerklärung ungemessen hoch.<sup>3</sup>

Ähnlich bedenklich stimmen die Ergebnisse von ANDERSON/KADOUS/KOONCE. Sie untersuchten, inwiefern sich Prüfer im Rahmen der Prüfungsplanung auf die Korrektheit einer Mandantenerklärung verlassen. Hierzu ließen sie insgesamt 113 Prüfer mit einer durchschnittlichen Prüfungserfahrung von 41 Monaten eine wesentliche Abweichung des Forderungsbestands und der Bruttogewinnmarge beurteilen.<sup>4</sup> Dabei erhielt die eine Prüfergruppe neben der Mandantenerklärung eine detaillierte Kalkulation, aus der hervorging, dass der gesamte Abweichungsbetrag auf den mitgeteilten Nicht-Fehlergrund zurückgeführt werden kann. Die andere Prüfergruppe erhielt hingegen ausschließlich die Mandantenerklärung ohne weiteren Nachweis.<sup>5</sup> Zusätzlich variierten die Autoren – ähnlich wie GLOVER/JIAMBALVO/KENNEDY – das jeweilige Prüfungsrisiko.<sup>6</sup> Im Ergebnis konnten ANDERSON/KADOUS/KOONCE feststellen, dass lediglich das Prüfungsrisiko

---

<sup>1</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 38-39 i.V.m. 41-42.

<sup>2</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 38-39.

<sup>3</sup> Vgl. GLOVER/JIAMBALVO/KENNEDY (2000), S. 41-42.

<sup>4</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 15-16.

<sup>5</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 16.

<sup>6</sup> Den Prüfern mit erhöhtem Prüfungsrisiko wurde u.a. mitgeteilt, dass der Mandant kurz vor einem Börsengang steht und dieses Jahr die Prüfungsgesellschaft gewechselt hat. Den anderen Prüfern wurde mitgeteilt, dass es sich um ein Familienunternehmen handelt, welches bereits seit mehreren Jahren von der gleichen Prüfungsgesellschaft geprüft wurde und bisher kaum Probleme verursacht hat. Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 16.

einen signifikanten Einfluss auf die Bereitschaft hatte, der Mandantenerklärung im Rahmen der weiteren Prüfungsplanung zu vertrauen. Auf einer Likert-Skala von eins bis zehn erzielten Prüfer, die einem hohen Prüfungsrisiko ausgesetzt waren, durchschnittlich 4,01 Punkte. Prüfer, die ein geringes Prüfungsrisiko zu berücksichtigen hatten, erreichten hingegen 5,09.<sup>1</sup> Keinen signifikanten Einfluss auf die Vertrauensbereitschaft hatte hingegen die Bereitstellung einer Berechnung, die die Verursachung der festgestellten Abweichung durch die Mandantenerklärung quantifizierte. Prüfer, denen lediglich die Mandantenerklärung zur Verfügung gestellt wurde, vertrauten der Erklärung des Mandanten in gleichem Maße wie Prüfer, die zusätzlich zur Mandantenerklärung eine Kalkulation erhielten (4,51 ggü. 4,62).<sup>2, 3</sup>

Zusammenfassend lassen die Erkenntnisse aus den Studien von GLOVER/JIAMBALVO/KENNEDY sowie ANDERSON/KADOUS/KOONCE ähnlich wie die Befunde zu den anderen Phasen der Durchführung von analytischen Prüfungshandlung erneut darauf schließen, dass von dem Erhalt einer inkorrekten Mandantenerklärung eine ernst zunehmende Gefahr ausgeht. Es ist folglich auch an dieser Stelle wiederholt darauf hinzuweisen, dass sich Prüfer dieser potenziellen Verzerrung durch eine Mandantenbefragung bewusst werden müssen, um einer möglichen (unbewussten) Beeinträchtigung der Prüfungsqualität entgegenwirken zu können.

---

<sup>1</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 22.

<sup>2</sup> Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 22.

<sup>3</sup> Die Autoren vermuteten u.a., dass diese Beobachtung möglicherweise darauf zurückzuführen sei, dass in der Praxis tätige Prüfer es für gewöhnlich nicht erwarten, auf eine derart umfassende Quantifizierung einer Mandantenerklärung in der (frühen) Planungsphase zurückgreifen zu können. Vgl. ANDERSON/KADOUS/KOONCE (2004), S. 25.



### 3.5.6.3 Tabellarische Darstellung der empirischen Befunde zur Phase der Prüfungsentscheidung

| Autor(en)                       | Teilnehmer |                                                           | Daten-<br>gewinnung | Daten-<br>auswertung       | Ergebnisse                                                                                                                                                                                                                                            | Signifikanz                            |
|---------------------------------|------------|-----------------------------------------------------------|---------------------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
|                                 | Anzahl     | Art                                                       |                     |                            |                                                                                                                                                                                                                                                       |                                        |
| BIGGS/MOCK/ WAT-<br>KINS (1988) | 4          | 2 Manager<br>Ø = 7 J.<br><br>2 Prüfungsleiter<br>Ø > 3 J. | Experiment          | Verbale Protokolle         | Prüfer erhöhen den Prüfungsumfang, wenn analytische Prüfungshandlungen eine wesentliche Abweichung identifizieren. Sie reduzieren den Prüfungsumfang jedoch nicht, wenn keine wesentliche Abweichung festgestellt wird (sog. asymmetrischer Einsatz). | k.A.                                   |
| COHEN/KIDA (1989)               | 96         | 50 Prüfungsleiter<br>k.A.                                 | Experiment          | t-Test ,<br>Varianzanalyse | Prüfer, die eine wesentliche Abweichung identifizierten, erhöhten den Prüfungsumfang signifikant im Vgl. zum ursprünglich vorgesehenen Prüfungsumfang und signifikant stärker als Prüfer, die keine wesentliche Abweichung identifizierten.           | 0,001 <sup>1</sup> /0,003 <sup>2</sup> |
|                                 |            | 46 Manager<br>k.A.                                        |                     | t-Test                     | Prüfer, die keine wesentliche Abweichung identifizierten, reduzierten den Prüfungsumfang nicht signifikant im Vgl. zum ursprünglich vorgesehenen Prüfungsumfang.                                                                                      | 0,608 <sup>3</sup> /0,188 <sup>4</sup> |
| COHEN (1994)                    | 84         | k.A.<br>Ø = 6 J.                                          | Experiment          | Varianzanalyse             | Prüfer, die eine wesentliche Abweichung identifizierten, erhöhten den Prüfungsumfang signifikant höher als Prüfer die keine wesentliche Abweichung feststellen konnten, und zwar unabhängig davon, ob eine Über- oder Unterbewertung vorlag.          | 0,021                                  |
|                                 |            |                                                           |                     | t-Test                     | Prüfer, die keine wesentliche Abweichung identifizierten, reduzierten den Prüfungsumfang nicht signifikant im Vgl. zum ursprünglich vorgesehenen Prüfungsum-                                                                                          | k.A.                                   |

<sup>1</sup> Signifikanzniveau des t-Tests.

<sup>2</sup> Signifikanzniveau der Varianzanalyse.

<sup>3</sup> Signifikanzniveau des t-Tests bei den Prüfungsleitern.

<sup>4</sup> Signifikanzniveau des t-Tests bei den Managern.

|                                  |     |                                                                       |            |                  |                                                                                                                                                                                                                                                                                                                          |       |
|----------------------------------|-----|-----------------------------------------------------------------------|------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|                                  |     |                                                                       |            |                  | fang.                                                                                                                                                                                                                                                                                                                    |       |
| FRA-SER/HATHERLY/LIN (1997)      | 366 | Partner<br>k.A.                                                       | Fragebogen | Häufigkeiten     | Mehr als die Hälfte der Befragten ist dazu bereit, den Prüfungsumfang zu reduzieren, falls das Ergebnis der analytischen Prüfungshandlung keine wesentliche Abweichung anzeigt.                                                                                                                                          | k.A.  |
| MULLIGAN/INGSTER (1999)          | 307 | Partner<br>Ø = 16,9 J.                                                | Fragebogen | Häufigkeiten     | 29% (35%) der Befragten reduzieren den Prüfungsumfang immer (häufig), falls das Ergebnis der analytischen Prüfungshandlung keine wesentliche Abweichung anzeigt. Allerdings lassen sie sich die Ergebnisse erst durch Ergebnisse anderer Prüfungshandlungen bestätigen, bevor sie den Prüfungsumfang reduzieren.         | k.A.  |
| LIN/FRASER (2003)                | 182 | 138 Partner<br>k.A.<br>39 Manager<br>k.A.<br>5 Prüfungsleiter<br>k.A. | Fragebogen | Häufigkeiten     | 50,5% der Befragten reduzieren den Prüfungsumfang nur dann, wenn die Ergebnisse durch andere Prüfungshandlungen bestätigt werden; 37,4% schränken den Umfang auch ohne weitere Prüfungsnachweise ein; 12,1% reduzieren den Umfang niemals, wenn das Ergebnis der analytischen Prüfungshandlung keine Abweichung anzeigt. | k.A.  |
| GLO-VER/JIAMBALVO/KENNEDY (2000) | 68  | k.A.<br>Ø = 3,1 J.                                                    | Experiment | Fisher Test      | Prüfer mit erhöhtem Prüfungsrisiko und low corroboration nahmen signifikant häufiger eine Erhöhung des Prüfungsumfangs vor als die übrigen Prüfer mit geringem Risiko und low corroboration, Prüfer mit geringem Risiko und high corroboration und Prüfer mit erhöhtem Risiko und high corroboration.                    | 0,007 |
|                                  |     |                                                                       |            | Chi-Quadrat-Test | Prüfer mit niedrigem Prüfungsrisiko und low corroboration führten nicht signifikant häufiger eine Erhöhung des Prüfungsumfangs durch als Prüfer mit niedrigem Prüfungsrisiko und high corroboration.                                                                                                                     | 0,610 |
|                                  |     |                                                                       |            | Chi-Quadrat-Test | Die Häufigkeit, mit der Prüfer mit erhöhtem Risiko und high corroboration den Prüfungsumfang erhöhten, weicht nicht signifikant von der Häufigkeit ab, mit der Prüfer mit niedrigem Prüfungsrisiko und high corroboration                                                                                                | 0,610 |

|                                        |     |                    |            |                |                                                                                                                                                    |      |
|----------------------------------------|-----|--------------------|------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------|------|
|                                        |     |                    |            |                | ration den Prüfungsumfang erhöhten.                                                                                                                |      |
| ANDER-<br>SON/KADOUS/<br>KOONCE (2003) | 113 | k.A.<br>Ø =41 Mon. | Experiment | Varianzanalyse | Prüfungsrisiko hat einen signifikanten Einfluss auf die Bereitschaft der Prüfer, einer Mandantenerklärung zu vertrauen.                            | 0,02 |
|                                        |     |                    |            | Varianzanalyse | Die Bereitstellung einer Quantifizierung hat keinen signifikanten Einfluss auf die Bereitschaft der Prüfer, einer Mandantenerklärung zu vertrauen. | 0,24 |

Tabelle 6: Ergebnisse der empirischen Prüfungsforschung zur Phase der Prüfungsentscheidung

### 3.6 Zwischenbetrachtung

Wie in Kap. 3.5 gezeigt wurde, hat sich die empirische Prüfungsforschung seit Anfang der achtziger Jahre umfassend mit der Durchführung von analytischen Prüfungshandlungen befasst. Dabei wurden zahlreiche interne und externe Einflussfaktoren festgestellt, die sich wesentlich auf die Effektivität und Effizienz von analytischen Prüfungshandlungen auswirken können. Im Mittelpunkt der Betrachtung standen dabei kognitive Verzerrungen, welche dadurch gekennzeichnet sind, dass sie nicht durch rationales Verhalten erklärt werden können. Handelte es sich zudem um prüfungsprozessimmanente Einflussfaktoren, so wurden diese zum Anlass genommen, um auf mögliche Missstände in der Prüfungspraxis aufmerksam zu machen. Ein weiterer wesentlicher Bestandteil der empirischen Forschungstätigkeit lag darin, potenzielle Mechanismen zu untersuchen, die einer effektivitäts- und effizienzmindernden Beeinflussung entgegenwirken können. Konnten sog. debiasing-Mechanismen identifiziert werden, so wurden auf dieser Basis Handlungsempfehlungen für die Prüfungspraxis ausgesprochen.

Obwohl die Forschungsbemühungen der letzten 25 Jahre sehr umfassend und vielseitig waren, nahm jedoch ein Aspekt eine dominierende Rolle ein, dem aufgrund seiner Manifestierung in den Prüfungsnormen und somit in der Vorgehensweise der Prüfungspraxis eine erhebliche Brisanz beizumessen ist. Hierbei handelt es sich um die Einholung einer Mandantenerklärung, falls analytische Prüfungshandlungen zur Aufdeckung einer wesentlichen unerwarteten Abweichung führen. Die potenzielle negative Beeinflussung der prüferischen Vorgehensweise durch den Erhalt einer Mandantenerklärung wurde dabei in allen Phasen des Urteilsbildungsprozesses untersucht. Im Rahmen der Problemrepräsentation konnte neben der Möglichkeit eines framing effects insbesondere nachgewiesen werden, dass der durch Erhalt einer inkorrekten Mandantenerklärung aufgebaute inadäquate Problemlösungsraum den Wechsel zu einer adäquaten Problemrepräsentation erheblich erschwert.<sup>1</sup> Der interference effect sowie der spreading effect stellen die beiden zentralen potenziellen Auswirkungen dar, die innerhalb der Hypothesengenerierung durch eine Mandantenerklärung ausgelöst werden können.<sup>2</sup> Wesentliche Aspekte in der Phase der Informationssuche sind die durch den Erhalt einer Mandantenerklärung zu befürchtende eingeschränkte Suchstrategie (truncated search strategy) sowie konfirmatorische Prozesse (confirmation bias).<sup>3</sup> Darüber hinaus kann eine Mandanten-

---

<sup>1</sup> Vgl. hierzu Kap. 3.5.2.2.1 sowie 3.5.2.3.

<sup>2</sup> Vgl. hierzu Kap. 3.5.3.3.2.1.2.1 und 3.5.3.3.2.1.2.2.

<sup>3</sup> Vgl. hierzu Kap. 3.5.4.4.3.

erklärung im Rahmen der Hypothesenbewertung dazu führen, dass der vererbten Hypothese eine höhere Eintrittswahrscheinlichkeit als anderen potenziellen plausiblen Abweichungsursachen zugeordnet wird.<sup>1</sup> Schließlich könnte sich eine Mandantenerklärung auch negativ auf die Phase der Prüfungsentscheidung auswirken.<sup>2</sup>

Die Frage nach der Adäquanz der Einholung einer Mandantenerklärung konnte trotz des enormen Forschungsaufwands nicht abschließend geklärt werden. Dieses wird insbesondere durch die z.T. widersprüchlichen Forschungsergebnisse dokumentiert. Darüber hinaus ist der bisherigen Prüfungsforschung eine einseitige und daher unvollständige Vorgehensweise zu attestieren.<sup>3</sup> Daher hat sich die vorliegende Arbeit zum Ziel gesetzt, einen weiteren Beitrag zur Klärung der (noch) offenen Fragen zu leisten. Hierzu wurde eine empirische Studie in Form eines Experiments konzipiert und durchgeführt, welche im nachfolgenden Kapitel ausführlich dargestellt wird.

---

<sup>1</sup> Vgl. hierzu Kap. 3.5.5.3, insbesondere jedoch 3.5.5.3.1.1.

<sup>2</sup> Vgl. hierzu Kap. 3.5.6.2.3.

<sup>3</sup> Vgl. hierzu ausführlich die unter Kap. 4.1 aufgeführten Erweiterungen der eigenen Studie gegenüber der bisherigen Prüfungsforschung.

## **4 Empirische Studie zur Durchführung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung**

### **4.1 Vorbemerkungen**

Wie in der Darstellung der empirischen Forschung zur Durchführung von analytischen Prüfungshandlungen mehrfach aufgezeigt wurde, stellt der mögliche Einfluss des Erhalts einer Abweichungserklärung des Mandanten auf die prüferische Urteilsbildung einen zentralen Gegenstand der bisherigen Forschungsbemühungen dar.<sup>1</sup> Die Tatsache, dass die Mandantenbefragung nicht nur die in der Prüfungsrealität praktizierte Vorgehensweise ist,<sup>2</sup> sondern darüber hinaus auch von den berufsständischen Prüfungsstandards empfohlen wird,<sup>3</sup> dokumentiert die Bedeutung der Identifikation der potenziellen Auswirkungen eines derartigen prüferischen Vorgehens. Daher wird sich auch die vorliegende Studie mit der Frage beschäftigen, inwieweit sich die Einholung einer Mandantenerklärung auf die weitere prüferische Urteilsbildung auswirkt.

Als Maßstab für die Feststellung eines möglichen Einflusses werden unterschiedliche Kriterien herangezogen. Zum einen beziehen sich diese Kriterien auf einzelne Phasen des in Kap. 3.3 vorgestellten Ablaufmodells analytischer Prüfungshandlungen.<sup>4</sup> Zum anderen sind Kriterien heranzuziehen, die sich von der phasenspezifischen Analyseebene lösen und beurteilen, ob der Erhalt einer Mandantenerklärung die Effektivität und Effizienz einer analytischen Prüfungshandlung im Gesamten beeinträchtigt.<sup>5</sup>

Die vorliegende Studie erweitert die bisherigen Forschungsaktivitäten in mehrfacher Hinsicht. Erstens fokussierte der überwiegende Teil der Studien einzelne Phasen des Ablaufmodells analytischer Prüfungshandlungen. So betrachteten gerade die frühen Studien ausschließlich die Phasen der Problemrepräsentation, der Hypothesengenerierung oder der Hypothesenbewertung.<sup>6</sup> Nur wenige Studien wandten sich von dieser

---

<sup>1</sup> Siehe hierzu auch die Ausführungen in Kap. 3.5.3.3.2.1.3, 3.5.4.4.3 sowie 3.5.5.3.1.1.

<sup>2</sup> So auch HIRST/KOONCE (1996), S. 462, 471.

<sup>3</sup> Vgl. IDW PS 312.27 sowie ISA 520.18.

<sup>4</sup> In Betracht kommen hierbei die Anzahl und die Art der intern generierten Hypothesen (Phase der Hypothesengenerierung), die Merkmale der angewandten Suchstrategie (Phase der Informationssuche), die Wahrscheinlichkeitsbeurteilung der generierten Hypothesen sowie die Auswahl einer Abweichungshypothese (Phase der Hypothesenbewertung).

<sup>5</sup> Effektivität bezieht sich dabei auf die Fehlerrückmeldungsfähigkeit der Prüfungshandlung. Effizienz berücksichtigt zusätzlich die mit den zur Fehlerrückmeldung vorgenommenen Prüfungsschritten in Verbindung stehenden angefallenen Kosten der Prüfungshandlung.

<sup>6</sup> Vgl. hierzu etwa die Studien von KAPLAN/RECKERS (1989), HEIMAN (1990), CHURCH/SCHNEIDER (1993), KOONCE (1992), NG/GREEN/SIMNETT (2001), ANDERSON/KOONCE (1995), ANDERSON/KOONCE (1998). Ausnahmen bilden wenige Studien, die den Zusammenhang zwischen der Hypothe-

„stückweisen“ Betrachtungsebene ab und orientierten sich vermehrt an einer ganzheitlichen Betrachtungsweise, die sowohl die einzelnen Phasen als auch das Gesamtergebnis einer analytischen Prüfungshandlung in den Mittelpunkt der Untersuchung rückt.<sup>1</sup> Eine derartige Betrachtung weist den immensen Vorteil auf, dass in Bezug auf phasenabhängige Verzerrungen zum einen festgestellt werden kann, ob diese im weiteren Verlauf der prüferischen Vorgehensweise fortgetragen werden und somit eine Bedrohung für die analytische Prüfungshandlung im Gesamten darstellen, oder ob diese Verzerrungen lediglich „stationären“ Charakter aufweisen, die in späteren oder wiederholten Phasen kompensiert werden können und keinen Einfluss auf das Gesamtergebnis der analytischen Prüfungshandlung ausüben. Da diese Frage von wesentlicher Bedeutung für die Beurteilung der Angemessenheit einer Mandantenbefragung ist, erscheinen weitere ganzheitliche Forschungsaktivitäten dringend notwendig.

Zweitens, den bisherigen Studien ist ausnahmslos vorzuwerfen, dass sie ausschließlich der Frage nachgingen, inwieweit sich eine Mandantenerklärung auswirkt, wenn die festgestellte Abweichung auf eine Fehlerursache zurückzuführen ist (im Folgenden als Fehler-Version bezeichnet).<sup>2</sup> Eine derartig einseitige Betrachtungsweise lässt jedoch unberücksichtigt, dass in der Prüfungspraxis möglicherweise ebenso häufig die Situation anzutreffen ist, in welcher der Prüfer eine Nicht-Fehlererklärung vom Mandanten erhält, die in Wahrheit für die festgestellte Abweichung verantwortlich ist (im Folgenden als Nicht-Fehler-Version bezeichnet). Um zu beantworten, welche Auswirkungen in der letztgenannten Situation zu erwarten sind, berücksichtigt die vorliegende Studie sowohl eine Fehler- als auch eine Nicht-Fehlerversion.

Drittens, die bisherigen Studien simulieren überwiegend Prüfungssituationen, in denen ein geringes Prüfungsrisiko vorliegt. Fraglich ist jedoch, ob die bisherigen Prüfungserkenntnisse auch in Prüfungssituationen zutreffen, die ein erhöhtes Prüfungsrisiko aufweisen.<sup>3</sup> Deshalb berücksichtigt die vorliegende Studie sowohl ein niedriges als auch ein erhöhtes Prüfungsrisiko.

---

sengenerierung und der anschließenden Informationssuche untersuchen (vgl. CHURCH (1991) sowie PETERSON/WONG-ON-WING (2000))

<sup>1</sup> Hierzu zählen die Studien von ASARE/WRIGHT/WRIGHT (1998), ASARE/TROMPETER/WRIGHT (2000), GREEN/TROTMAN (2003), ASARE/WRIGHT (2003), GREEN (2004a), GREEN (2004b).

<sup>2</sup> Überwiegend handelte es sich hierbei um eine inkorrekte Nicht-Fehlererklärung des Mandanten. Eine Ausnahme bildet die Studie von BEDARD/BIGGS (1991b), welche die Auswirkung der Mitteilung einer korrekten Fehlerursache durch den Mandanten untersucht.

<sup>3</sup> Eine Empfehlung zur Integration des Prüfungsrisikos spricht auch GREEN (2004a), S. 388 aus.

Viertens, bisher existieren nur zwei Studien, die den Einfluss von branchenspezifischer Erfahrung auf die Effektivität und Effizienz von analytischen Prüfungshandlungen untersuchen.<sup>1</sup> Da in der Prüfungspraxis eine zunehmende Branchenspezialisierung von Prüfungsleistungen zu verzeichnen ist, kann eine möglichst hohe Praxisrelevanz und Akzeptanz von Forschungsergebnissen durch Berücksichtigung der Branchenerfahrung erreicht werden.<sup>2</sup> Um die Erkenntnisse über die Auswirkungen von Branchenspezifität voranzutreiben, wird in der vorliegenden Studie u.a. zwischen Branchenspezialisten einerseits (im Folgenden als Spezialisten bezeichnet) und Nicht-Branchenspezialisten andererseits (im Folgenden als Nicht-Spezialisten bezeichnet) differenziert.

Neben diesen inhaltlichen Erweiterungen der bisherigen Prüfungsforschung kann ein weiterer Nutzen dieser Studie auch darin gesehen werden, dass die bisherigen umfangreichen Forschungserkenntnisse, die überwiegend dem angelsächsischen Sprachraum entstammen, einer Rauminvarianz-Überprüfung unterzogen werden.<sup>3</sup> Insofern ist es von besonderem Interesse festzustellen, ob die bisherigen Erkenntnisse auch für den nationalen Bereich Gültigkeit besitzen.

## **4.2 Formulierung der Hypothesen und Forschungsfragen**

### **4.2.1 Der Einfluss des Erhalts einer Mandantenerklärung auf die interne Hypothesengenerierung**

Ein zentraler Gegenstand der bisherigen Forschungsbemühungen ist in der Untersuchung der Auswirkung des Erhalts einer Nicht-Fehlererklärung durch den Mandanten auf die anschließende interne Hypothesengenerierung zu sehen. Mögliche Folgen einer „vererbten“ Nicht-Fehlererklärung stellen die mangelnde Fähigkeit zur anschließenden Generierung von weiteren Fehlerhypothesen (sog. interference effect) und die vermehrte

---

<sup>1</sup> Dabei handelt es sich um die Studien von BEDARD/BIGGS (1991b) und GREEN (2004b). Zudem ist dem Autor nur die Studie von WRIGHT/WRIGHT (1997) bekannt, die den Einfluss der branchenspezifischen Erfahrung auf die Hypothesengenerierung untersucht.

<sup>2</sup> Für die Relevanz der Berücksichtigung von branchenspezifischer Erfahrung spricht sich auch GREEN (2004b) aus, die den Einfluss branchenspezialisierter Prüfungsleistungen auf die Effektivität und Effizienz von analytischen Prüfungshandlungen untersucht. Im Ergebnis stellt GREEN fest, dass Branchenspezialisten wesentlich häufiger den tatsächlichen Grund der festgestellten Abweichung identifizieren und dabei weniger Prüfungszeit in Anspruch nehmen.

<sup>3</sup> Vgl. hierzu ausführlich die Kritik an der Aufstellung nomologischer Hypothesen im Rahmen des empirischen Forschungsprogramms in Kap. 3.2.1.2.1.2.



Generierung von weiteren Nicht-Fehlerhypothesen (sog. spreading effect) dar.<sup>1</sup> Die Existenz des spreading effect konnte in der empirischen Prüfungsforschung bereits mehrfach nachgewiesen werden.<sup>2</sup> In Bezug auf den interference effect lieferten die bisherigen Forschungsbemühungen jedoch widersprüchliche Ergebnisse.<sup>3</sup> Einerseits konnte eine mangelnde Fähigkeit zur Generierung von weiteren Fehlerhypothesen festgestellt werden, andererseits ließ sich jedoch kein Einfluss einer Nicht-Fehlererklärung auf die Anzahl der selbstgenerierten Fehlerhypothesen nachweisen.<sup>4</sup> Diese konfliktären Ergebnisse könnten möglicherweise darauf zurückgeführt werden, dass einige Studien aufgrund ihres zugrunde liegenden Forschungsdesigns lediglich dazu in der Lage waren, ihre Ergebnisse auf Basis der anfänglichen Hypothesengenerierung herzuleiten<sup>5</sup>, andere Studien hingegen die gesamte (d.h. sowohl die anfängliche als auch die anschließende<sup>6</sup>) Hypothesengenerierung zu berücksichtigen.<sup>7</sup>

Um den interference und den spreading effect einer genaueren Überprüfung zu unterziehen, soll der Zusammenhang zwischen dem Erhalt einer Nicht-Fehlererklärung durch den Mandanten und der internen Hypothesengenerierung anhand der unten stehenden Hypothesen genauer untersucht werden. Um dabei einen möglichst hohen Erkenntnisgewinn zu erlangen, ist jeweils zwischen der anfänglichen und der gesamten Hypothesengenerierung zu differenzieren:

---

<sup>1</sup> Für den theoretischen Hintergrund sowie die bisherigen Untersuchungserkenntnisse zum interference effect einerseits und zum spreading effect andererseits soll auf die Ausführungen in Kap. 3.5.3.3.1.2 i.V.m. 3.5.3.3.1.2 verwiesen werden.

<sup>2</sup> Siehe hierzu ausführlich Kap. 3.5.3.3.2.1.2.2.

<sup>3</sup> Siehe hierzu ausführlich Kap. 3.5.3.3.2.1.2.1.

<sup>4</sup> Einen Nachweis für die Existenz eines interference effects erbrachte HEIMAN (1990), S. 880, 883. Zudem konnten BIERSTAKER/BEDARD/BIGGS (1999), S. 32 eine eingeschränkte Hypothesengenerierung (sog. truncated hypothesis generation) feststellen, die durch den Erhalt einer Mandantenerklärung verursacht wurde. Dieses Phänomen weist eine unverkennbare Ähnlichkeit zum interference effect auf. Keinen interference effect bei Erhalt einer Mandantenerklärung konnten indes NG/GREEN/SIMNETT (2001) sowie GREEN (2004a) erkennen.

<sup>5</sup> Vgl. etwa HEIMAN (1990).

<sup>6</sup> Zur Unterscheidung zwischen der anfänglichen und anschließenden Hypothesengenerierung soll an dieser Stelle auf die Ausführungen in Kap. 3.5.3.2 verwiesen werden.

<sup>7</sup> Vgl. etwa GREEN (2004a).

- H<sub>1.1</sub>:*** *Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.*
- H<sub>1.2</sub>:*** *Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.*
- H<sub>1.3</sub>:*** *Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.*
- H<sub>1.4</sub>:*** *Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.*

Aufgrund der in der Praxis zunehmenden Branchenspezialisierung von Prüfungsdienstleistungen hat sich die empirische Prüfungsforschung verstärkt dem Einfluss branchenspezifischer Erfahrung auf die Vorgehensweise des Prüfers gewidmet.<sup>1</sup> In Bezug auf die Durchführung von analytischen Prüfungshandlungen konnte festgestellt werden, dass sich die Hypothesengenerierung von Spezialisten einerseits und Nicht-Spezialisten andererseits nicht signifikant voneinander unterscheidet.<sup>2</sup> Allerdings wurde nicht untersucht, inwiefern sich der Erhalt einer Mandantenerklärung auf die Hypothesengenerierung auswirkt. Daher soll anhand folgender Hypothesen untersucht werden, ob ein Zusammenhang zwischen der branchenspezifischen Erfahrung der Prüfer und dem Einfluss einer Mandantenerklärung auf die Hypothesengenerierung besteht:

---

<sup>1</sup> Vgl. stellvertretend SOLOMON/SHIELDS/WHITTINGTON (1999).

<sup>2</sup> Vgl. GREEN (2004b), S. 12.

- H<sub>1.1.1</sub>***: Nicht-Spezialisten, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.2.1</sub>***: Nicht-Spezialisten, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.1.2</sub>***: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.2.2</sub>***: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.3.1</sub>***: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.4.1</sub>***: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.3.2</sub>***: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.
- H<sub>1.4.2</sub>***: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.

Es ist weiterhin fraglich, ob möglicherweise ein Zusammenhang zwischen der Beeinflussung der Hypothesengenerierung durch die Vererbung einer Nicht-Fehlererklärung und dem zugrunde liegenden Prüfungsrisiko besteht. Denkbar wäre etwa, dass die o.a. potenziellen Verzerrungen lediglich bei geringem Prüfungsrisiko vorliegen, womit er-

höhtem Prüfungsrisiko mithin eine Debiasing-Funktion zu bescheinigen wäre. Da diese Fragen in der empirischen Prüfungsforschung bisher außer Acht gelassen wurden, sollen die aufgeführten Hypothesen anhand der folgenden explorativ ausgerichteten Forschungsfragen ergänzt werden:

**Forschungsfrage 1.1:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die anfängliche Hypothesengenerierung und dem Prüfungsrisiko?*

**Forschungsfrage 1.2:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die gesamte Hypothesengenerierung und dem Prüfungsrisiko?*

#### **4.2.2 Der Einfluss des Erhalts einer Mandantenerklärung auf die Informationssuche**

##### **4.2.2.1 Der Einfluss auf die angewandte Suchstrategie**

Nachdem der Prüfer potenzielle Hypothesen für eine festgestellte Abweichung generiert hat, obliegt es ihm, Prüfungsnachweise einzuholen, um seine Vermutungen zu bestätigen oder zu widerlegen.<sup>1</sup> In diesem Kontext beschäftigte sich ein Teil der empirischen Prüfungsforschung mit dem möglichen Einfluss, den der Erhalt einer Mandantenerklärung auf das prüferische Suchverhalten ausübt. Im Mittelpunkt standen dabei die Auswirkungen auf den Umfang, der Breite und die Tiefe der Informationssuche.<sup>2</sup> Insbesondere wurde vermutet, dass die „vererbte“ Hypothese zur Anwendung einer positiven Suchstrategie führe, welche dadurch gekennzeichnet ist, dass sich der Prüfer überwiegend auf Informationen konzentriert, die die Korrektheit der Mandantenerklärung bestätigen. Die Fokussierung auf die Mandantenerklärung führe zu einer Vernachlässigung weiterer potenzieller Abweichungsursachen, was einer ausgewogenen bzw. breiten Suchstrategie entgegenstehe.

Die empirische Prüfungsforschung brachte auch hier unterschiedliche Befunde hervor. Zum einen konnten hinsichtlich der Sucheigenschaften keine signifikanten Unterschiede

---

<sup>1</sup> Für eine ausführliche Darstellung der Eigenschaften der Informationssuche sei an dieser Stelle auf Kap. 3.5.4.3 verwiesen.

<sup>2</sup> In einem unmittelbaren Zusammenhang zu den Eigenschaften der Informationssuche stehen die positive, negative und ausgewogene Suchstrategie. Siehe hierzu ausführlich Kap. 3.5.4.1.

zwischen Prüfern mit und ohne Mandantenerklärung festgestellt werden.<sup>1</sup> Zum anderen konnte nachgewiesen werden, dass Prüfer mit Mandantenerklärung wesentlich häufiger auf Informationen zurückgriffen, die auf die Überprüfung der Mandantenerklärung gerichtet waren, und ihr Suchverhalten somit tiefer ausfiel als bei Prüfern ohne Mandantenerklärung.<sup>2</sup>

Um die Interdependenz zwischen dem Erhalt einer Mandantenerklärung und den Eigenschaften des prüferischen Suchverhaltens genauer zu untersuchen, soll die Informationssuche anhand folgender Hypothesen überprüft werden:

***H<sub>II.1</sub>:*** *Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Prüfer ohne Mandantenerklärung.*

***H<sub>II.2</sub>:*** *Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Prüfer ohne Mandantenerklärung.*

In der empirischen Prüfungsforschung wurde bisher untersucht, inwiefern sich branchenspezifische Erfahrung auf die Suchstrategie auswirkt. Dabei wurde festgestellt, dass Branchenspezialisten zur Identifikation einer Abweichungsursache weniger Tests (Umfang) benötigten und dabei eine geringere Anzahl potenzieller Hypothesen untersuchten (Breite) als Nicht-Spezialisten.<sup>3</sup> Nicht betrachtet wurde hingegen, inwiefern sich der Erhalt einer Mandantenerklärung auf die Suchstrategie auswirkt. Daher soll anhand folgender Hypothesen untersucht werden, inwiefern ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Suchstrategie und die branchenspezifische Erfahrung besteht:

---

<sup>1</sup> Vgl. GREEN (2004a); ASARE/WRIGHT/WRIGHT (1998).

<sup>2</sup> Vgl. ANDERSON/KOONCE (1995) sowie BIERSTAKER/BEDARD/BIGGS (1999). Auch ASARE/WRIGHT/WRIGHT (1998) stellten in ihrer Untersuchung fest, dass Prüfer mit Mandantenerklärung wesentlich häufiger Informationen einholten, die auf die Beurteilung der Mandantenerklärung gerichtet waren.

<sup>3</sup> Vgl. GREEN (2004b), S. 13.

**H<sub>II.1.1</sub>:** *Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Nicht-Spezialisten ohne Mandantenerklärung.*

**H<sub>II.2.1</sub>:** *Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Nicht-Spezialisten ohne Mandantenerklärung.*

**H<sub>II.1.2</sub>:** *Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Spezialisten ohne Mandantenerklärung.*

**H<sub>II.2.2</sub>:** *Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Spezialisten ohne Mandantenerklärung.*

Da in der Prüfungsforschung bisher noch keine Erkenntnisse hinsichtlich eines möglichen Zusammenhangs zwischen dem Einfluss einer Mandantenerklärung auf die Suchstrategie und dem Prüfungsrisiko vorliegen, ist die Untersuchung um folgende explorative Forschungsfrage zu ergänzen:

**Forschungsfrage II:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die angewandte Suchstrategie und dem Prüfungsrisiko?*

#### **4.2.2.2 Der Einfluss auf die Interpretation der eingeholten Prüfungsnachweise**

Neben der angewandten Suchstrategie stellt auch die Interpretation der erlangten Informationen einen möglichen Gegenstand dar, welcher durch den Erhalt einer Mandantenerklärung nachhaltig beeinflusst werden kann.<sup>1</sup> So könnte der Prüfer etwa Informationen, welche die Korrektheit der Mandantenerklärung widerlegen oder zumindest be-

---

<sup>1</sup> Eine ausführliche Darstellung sog. konfirmatorischer Prozesse findet sich unter Kap. 3.5.4.2.1.

zweifeln lassen, oder Informationen, die auf die Gültigkeit einer anderen Abweichungshypothese hindeuten, (unbewusst) fehlinterpretieren oder einfach übergehen (sog. „explaining away“ disconfirming evidence). Darüber hinaus könnte der Erhalt einer Mandantenerklärung auch dazu führen, dass die Informationssuche des Prüfers überwiegend auf die Bestätigung der vererbten Hypothese ausgerichtet ist. Derartige konfirmatorische Prozesse konnten bei selbstgenerierten Hypothesen bereits mehrfach nachgewiesen werden.<sup>1</sup> Einige Untersuchungen lieferten Hinweise darauf, dass konfirmatorische Prozesse auch bei „vererbten“ Hypothesen vorliegen.<sup>2</sup> Daher soll der Einfluss des Erhalts einer Mandantenerklärung auf die Interpretation der erlangten Prüfungsnachweise anhand folgender Hypothesen genauer untersucht werden:

- H<sub>III.1</sub>: Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Prüfer ohne Mandantenerklärung.*
- H<sub>III.2</sub>: Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Prüfer ohne Mandantenerklärung.*

In der empirischen Prüfungsforschung konnte nachgewiesen werden, dass Spezialisten Informationen häufiger richtig interpretieren als Nicht-Spezialisten.<sup>3</sup> Allerdings liegen bisher keine Studien vor, die untersuchen, ob eine Interdependenz zwischen der branchenspezifischen Erfahrung und dem Einfluss einer Mandantenerklärung auf die Interpretation von Informationen besteht. Daher sollen folgende Hypothesen aufgenommen werden:

---

<sup>1</sup> Vgl. KAPLAN/RECKERS (1989); BEDARD/BIGGS (1991a) sowie HEIMAN-HOFFMAN/MOSER/JOSEPH (1995).

<sup>2</sup> Vgl. ANDERSON/KOONCE (1995); BIERSTAKER/BEDARD/BIGGS (1999).

<sup>3</sup> Vgl. GREEN (2004b), S. 14.

*H<sub>III.1.1</sub>: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Nicht-Spezialisten ohne Mandantenerklärung.*

*H<sub>III.2.1</sub>: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Nicht-Spezialisten ohne Mandantenerklärung.*

*H<sub>III.1.2</sub>: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Spezialisten ohne Mandantenerklärung.*

*H<sub>III.2.2</sub>: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Spezialisten ohne Mandantenerklärung.*

Folgende explorativ ausgerichtete Forschungsfrage dient der Feststellung eines möglichen Zusammenhangs zwischen dem Einfluss einer Mandantenerklärung auf die Interpretation von Informationen und dem Prüfungsrisiko:

**Forschungsfrage III:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Interpretation von Informationen und dem Prüfungsrisiko?*

#### **4.2.3 Der Einfluss des Erhalts einer Mandantenerklärung auf die Hypothesenbewertung**

Nachdem der Prüfer die Informationssuche abgeschlossen hat, obliegt es ihm, die generierten Hypothesen im Lichte der erlangten Prüfungsnachweise zu beurteilen. Dabei werden die generierten Hypothesen vor dem Hintergrund der erlangten Informationen einer Wahrscheinlichkeitsbeurteilung unterzogen und die Hypothese mit der höchsten Wahrscheinlichkeit selektiert. Der Erhalt einer Mandantenerklärung könnte dabei be-



wirken, dass die Wahrscheinlichkeit der Mandantenerklärung höher eingeschätzt und folglich häufiger als der vermutete Grund für die festgestellte Abweichung selektiert wird.<sup>1</sup> In der empirischen Prüfungsforschung wurde dieser Zusammenhang bereits mehrfach untersucht. Allerdings stellten lediglich zwei Studien die Auswahlhäufigkeit der Mandantenerklärung der entsprechenden Auswahlhäufigkeit von Prüfern ohne Mandantenerklärung gegenüber.<sup>2</sup> Hierbei wurden unterschiedliche Ergebnisse festgestellt.<sup>3</sup> Daher soll der Zusammenhang zwischen dem Erhalt einer Mandantenerklärung und deren Auswahl als wahrscheinlichste Abweichungsursache anhand folgender Hypothese genauer untersucht werden:

*H<sub>IV.1</sub>: Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Prüfer ohne Mandantenerklärung.*

Da in der empirischen Prüfungsforschung bisher keine Erkenntnisse vorliegen, die den Einfluss des Erhalts einer Abweichungserklärung des Mandanten auf die anschließende Bewertung derselben in Abhängigkeit von der branchenspezifischen Erfahrung der Prüfer eruieren, soll dieser Zusammenhang anhand folgender Hypothesen untersucht werden:

*H<sub>IV.1.1</sub>: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Nicht-Spezialisten ohne Mandantenerklärung.*

*H<sub>IV.1.2</sub>: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Spezialisten ohne Mandantenerklärung.*

---

<sup>1</sup> Für eine weitergehende Erörterung des potenziellen Einflusses des Erhalts einer Mandantenerklärung auf die Hypothesenbewertung sei an dieser Stelle auf Kap. 3.5.5.3.1.1 i.V.m. 3.5.5.1 verwiesen.

<sup>2</sup> Eine Studie, die lediglich die Auswahlhäufigkeit von Prüfern mit Mandantenerklärung untersuchte, führten etwa GREEN/TROTMAN (2003) durch.

<sup>3</sup> GREEN (2004a) etwa stellte fest, dass 17 von 42 Prüfern mit inkorrekt erklärter Mandantenerklärung diese auch als wahrscheinlichste Abweichungsursache selektierten, wohingegen nicht einer von 20 Prüfern ohne Mandantenerklärung den entsprechenden Grund auswählte. Weniger eindeutig fielen die Ergebnisse in der Untersuchung von ASARE/WRIGHT/WRIGHT (1998) aus (67,5% im Vergleich zu 43,7%).

Ferner dient folgende explorativ ausgerichtete Forschungsfrage der Untersuchung, inwiefern die Bewertung einer Mandantenerklärung durch das Prüfungsrisiko beeinflusst wird:

**Forschungsfrage IV:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die anschließende Bewertung dieser Erklärung und dem Prüfungsrisiko?*

#### **4.2.4 Der Einfluss des Erhalts einer Mandantenerklärung auf die Erfolgsquote**

Berücksichtigt man die in den bisher hergeleiteten Hypothesen unterstellten Auswirkungen des Erhalts einer Mandantenerklärung, so ist zwangsläufig zu vermuten, dass der Erhalt einer Mandantenerklärung auch nachhaltig die Erfolgsquote<sup>1</sup> des Prüfers beeinflusst. Prüfer ohne Mandantenerklärung – und ohne die daraus resultierenden Verzerrungen im Rahmen der Hypothesengenerierung, der Informationssuche und der Hypothesenbewertung – sollten wesentlich häufiger den tatsächlichen Grund einer festgestellten Abweichungsursache identifizieren als Prüfer mit Mandantenerklärung. Allerdings wurde diese Hypothese im Rahmen der bisherigen Forschungsaktivitäten erst zweimal untersucht und dabei nicht bestätigt.<sup>2</sup> Um den vermuteten Zusammenhang zwischen dem Erhalt einer Mandantenerklärung und der Erfolgsquote einer erneuten Untersuchung zu unterziehen, sollen abschließend folgende Hypothesen berücksichtigt werden:

*H<sub>V.1</sub>: Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus als Prüfer ohne Mandantenerklärung.*

In der empirischen Prüfungsforschung konnte bisher nachgewiesen werden, dass Spezialisten im Rahmen analytischer Prüfungshandlungen signifikant erfolgreicher waren als

---

<sup>1</sup> Unter der *Erfolgsquote* wird an dieser Stelle die Fähigkeit des Prüfers verstanden, den tatsächlichen Abweichungsgrund festzustellen. Da es sich bei der Erfolgsquote sowohl um die korrekte Identifikation einer Fehler- als auch Nicht-Fehlerursache handeln kann, ist die Erfolgsquote nicht mit der *Effektivität* gleichzusetzen, da Letztere per definitionem ausschließlich die Fehleraufdeckungsfähigkeit betrachtet.

<sup>2</sup> Vgl. BEDARD/BIGGS (1991b) sowie GREEN (2004a).

Nicht-Spezialisten.<sup>1</sup> Nicht untersucht wurde hingegen, inwiefern ein Zusammenhang zwischen branchenspezifischer Erfahrung und der Erfolgsquote bei Erhalt einer Mandantenerklärung besteht. Dieses soll anhand folgender Hypothesen untersucht werden:

*H<sub>V.1.1</sub>: Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus, als Nicht-Spezialisten ohne Mandantenerklärung.*

*H<sub>V.1.2</sub>: Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus als Spezialisten ohne Mandantenerklärung.*

Weiterhin soll mittels folgender Forschungsfrage untersucht werden, ob der Einfluss des Erhalts einer Mandantenerklärung auf die Erfolgsquote in Relation zum vorliegenden Prüfungsrisiko zu betrachten ist:

**Forschungsfrage<sub>v</sub>:** *Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Erfolgsquote und dem Prüfungsrisiko?*

---

<sup>1</sup> Vgl. GREEN (2004b), S. 15.

## 4.3 Forschungsdesign

### 4.3.1 Überblick über die ausgewählte Forschungsmethode und die Untersuchungsteilnehmer

Für die Untersuchung der zuvor aufgestellten Hypothesen wurde ein Experiment konzipiert, in welchem die teilnehmenden Probanden den Grund einer mittels analytischer Prüfungshandlung festgestellten wesentlichen Abweichung bei einem Mandanten der Industriebranche identifizieren sollten. Das Experiment umfasst drei Kontrollvariablen mit jeweils zwei Ausprägungen.<sup>1</sup> Die erste Variable stellt die Mandantenbefragung dar. Die Probanden wurden in zwei Gruppen klassifiziert, wobei der einen Gruppe eine Mandantenerklärung für die zu untersuchende Abweichung mitgeteilt wurde und der anderen Gruppe hingegen keine Mandantenerklärung bereitgestellt wurde. Die zweite Variable ist der tatsächliche Grund für die festgestellte Abweichung. Hierbei wurde zwischen einer Fehler- und einer Nicht-Fehler-Version differenziert. Bei erstgenannter Version wurde die festgestellte Abweichung durch einen Fehler in der Rechnungslegung des Mandanten verursacht. Hingegen wurde die Abweichung im Rahmen der Nicht-Fehler-Version durch eine Änderung in der Unternehmensumwelt hervorgerufen. Die dritte Variable umfasst das zugrunde liegende Prüfungsrisiko, wobei zwischen einem niedrigen und einem erhöhten Prüfungsrisiko zu unterscheiden ist. Insgesamt ergibt sich folglich eine 2×2×2-Untersuchungsmatrix, aus der acht unterschiedliche Fallvarianten (Zellen) hervorgehen.<sup>2</sup>

Teilnehmer der Untersuchung waren 87 in der Praxis tätige Prüfer der Big Four Prüfungsgesellschaften, welche zufällig einer der acht möglichen Fallvarianten zugewiesen wurden. Die Teilnehmer wiesen eine durchschnittliche Prüfungserfahrung von 56,81

---

<sup>1</sup> Eine ausführliche Darstellung von Laborexperimenten sowie deren Vor- und Nachteile findet sich unter Kap. 3.4.2.2.

<sup>2</sup> Allerdings ist an dieser Stelle darauf hinzuweisen, dass aufgrund der eingeschränkten Teilnehmerzahl z.T. eine Aggregation von Fallvarianten durchgeführt werden musste. Eine Aggregation war dann zwingend notwendig, wenn die Ergebnisse separat für Spezialisten und Nicht-Spezialisten dargestellt wurden. Aufgrund der zu geringen Teilnehmerzahl in den Zellen wurde hier jeweils auf die Differenzierung zwischen geringem und erhöhtem Prüfungsrisiko verzichtet, womit sich diesbezüglich eine 2×2 Untersuchungsmatrix ergibt (siehe Kap. 4.4.3.3.2, 4.4.4.2.2, 4.4.4.3.3, 4.4.5.2, 4.4.6.2). Keine Aggregation der unterschiedlichen Risikoausprägungen war hingegen bei der Untersuchung der anfänglichen Hypothesengenerierung der Spezialisten und Nicht-Spezialisten notwendig (siehe Kap. 4.4.3.2.2), da hier eine Zusammenfassung der Fehler- und Nicht-Fehler-Version vorgenommen wurde (siehe hierzu die Begründung in Kap. 4.4.3.1). Im Gegensatz zu den Ergebnissen mit Berücksichtigung der branchenspezifischen Erfahrung konnte bei den Ergebnissen ohne Berücksichtigung branchenspezifischer Erfahrung (siehe Kap. 4.4.3.3.1, 4.4.4.2.1, 4.4.4.3.2, 4.4.5.1, 4.4.6.1) eine Differenzierung zwischen niedrigem und erhöhtem Prüfungsrisiko beibehalten werden, da hier durch den Wegfall der Klassifikation nach Spezialisten und Nicht-Spezialisten eine ausreichende Teilnehmerzahl in den jeweiligen Zellen vorlag.

Monaten mit einer Standardabweichung von 49,02 Monaten auf. 46 der Probanden bezeichneten sich im Rahmen einer Selbsteinschätzung als Spezialisten der Industriebranche (im Folgenden als Spezialisten bezeichnet). Die übrigen Probanden schätzten sich hingegen nicht als Spezialisten dieser Branche ein (im Folgenden als Nicht-Spezialisten bezeichnet).<sup>1</sup> Die Selbsteinschätzung der Teilnehmer konnte durch die weiteren Angaben der Teilnehmer gestützt werden. Wie in Tabelle 7 zu erkennen ist, weisen die Spezialisten signifikant mehr branchenspezifische Erfahrung<sup>2</sup> als Nicht-Spezialisten auf (43,7 ggü. 17,6 Monaten), verfügen über einen signifikant höheren prozentualen Anteil an Mandanten mit wesentlichen Vorratsbeständen (60,8% ggü. 35,9%) und nahmen an wesentlich mehr Vorratsprüfungen teil (47,1% ggü. 35,9%).<sup>3</sup> Im Gegensatz zu diesen speziellen Erfahrungsarten bestehen zwischen den Spezialisten und Nicht-Spezialisten jedoch keine signifikanten Unterschiede hinsichtlich der allgemeinen Prüfungserfahrung (59,8 Monate im Vgl. zu 53,4 Monate).

|                                                                                          | <b>Spezialisten<br/>n=46</b> | <b>Nicht-Spezialisten<br/>n=41</b> | <b>Signifikanz<sup>4</sup><br/>(2-seitig)</b> |
|------------------------------------------------------------------------------------------|------------------------------|------------------------------------|-----------------------------------------------|
| Prüfungserfahrung [Monate]                                                               | 59,8                         | 53,4                               | 0,316                                         |
| Branchenspezifische Erfahrung [Monate]                                                   | 43,7                         | 17,6                               | 0,000                                         |
| Anteil der Mandanten mit wesentlichen Vorräten [%]                                       | 60,8                         | 35,9                               | 0,000                                         |
| Anteil der Prüfungsaufträge, bei denen an der Prüfung der Vorräte teilgenommen wurde [%] | 47,1                         | 27,1                               | 0,001                                         |

Tabelle 7: Umfang der durchschnittlichen Erfahrungsarten der Spezialisten und Nicht-Spezialisten

<sup>1</sup> Die Frage, ob sich der jeweilige Teilnehmer als Spezialist der Industriebranche einschätzt, wurde neben diversen anderen Fragen im Rahmen der demographischen Fragen gestellt, welche nach der Fallbearbeitung zu beantworten waren.

<sup>2</sup> Der Umfang der branchenspezifischen Erfahrung wurden durch das Produkt der bisherigen Prüfungsjahre und des prozentualen Anteils der gesamten Prüfungsaufträge, die der Industriebranche zuzurechnen sind, ermittelt. Beide Angaben wurden im Rahmen der demographischen Fragen getätigt.

<sup>3</sup> Der Anteil der Mandanten mit wesentlichen Vorräten sowie der Anteil der Prüfungsaufträge, bei denen an der Prüfung der Vorratsbestände mitgewirkt wurde, waren ebenfalls im Rahmen der demographischen Fragen anzugeben. Diese beiden Erfahrungsarten wurden in der Studie von BEDARD/BIGGS (1991b) als Maßstab für die sog. domänenspezifische Erfahrung der Probanden herangezogen. Da die Unterschiede des Umfangs an domänenspezifischer Erfahrung zwischen den Spezialisten und Nicht-Spezialisten höchst signifikant sind ( $p = 0,000$  bzw.  $p = 0,001$ ), kann die in der vorliegenden Studie vorgenommene Klassifizierung nach der Branchenspezifität auch als Unterteilung zwischen domänen-erfahrenen und –unerfahrenen Prüfern aufgefasst werden.

<sup>4</sup> Da die Daten zu den jeweiligen Erfahrungsarten nicht normalverteilt und die Voraussetzungen zur Durchführung von parametrischen Testverfahren somit nicht erfüllt sind, wurde von der Durchführung des t-Tests zugunsten des nicht-parametrischen Mann-Whitney-U-Tests abgesehen.

Ein Großteil der Probanden (72,4%) wurde über unterschiedliche Kontaktpersonen der jeweiligen Prüfungsgesellschaften akquiriert. Die Kontaktpersonen waren ausschließlich Mitarbeiter mit leitender Funktion, welche in ihrem Namen eine Sammelmail an die in ihrem Verantwortungsbereich tätigen Mitarbeiter verschickten.<sup>1</sup> Die Sammelmail enthielt einen Hinweis auf eine mögliche Untersuchungsteilnahme sowie ein Anschreiben, welches interessierten Mitarbeitern nähere Informationen zur weiteren Vorgehensweise zur Verfügung stellte.<sup>2</sup> Die Namen der angeschriebenen Mitarbeiter sowie ihre Gesamtzahl wurden von den Kontaktpersonen aus Gründen der Anonymität nicht bekannt gegeben, so dass keine Aussagen über eine Rücklaufquote getroffen werden können. Um die Teilnehmerzahl zu erhöhen, wurde zusätzlich eine Stichprobe von 300 Wirtschaftsprüfern aus dem Wirtschaftsprüferverzeichnis online gezogen.<sup>3, 4</sup> Den ausgewählten Wirtschaftsprüfern wurde ein Anschreiben zugesandt, in dem auf eine mögliche Teilnahme hingewiesen wurde und die notwendigen Informationen zur weiteren Vorgehensweise bereitgestellt wurden.<sup>5</sup> Aus dieser Stichprobe nahmen 24 Prüfer teil. Dies entspricht einer Rücklaufquote von 8%.

#### **4.3.2 Beschreibung des Falls und des Ablaufs der Fallbearbeitung**

Der zu bearbeitende Fall orientiert sich weitestgehend an der Aufgabenstellung aus den Studien von ASARE/WRIGHT (2003), GREEN/TROTMAN (2003) sowie GREEN (2004a, 2004b). Dabei wurde die Aufgabenstellung an die Erfordernisse einer handelsrechtlichen Jahresabschlussprüfung angepasst.<sup>6</sup> Die teilnehmenden Prüfer wurden in dem Anschreiben auf eine Internetseite verwiesen, auf der sie einerseits die notwendigen Fallunterlagen herunterladen und andererseits die Fallstudie online bearbeiten konnten. Zum Herunterladen der Unterlagen benötigten die Teilnehmer ein Passwort, welches ebenfalls dem Anschreiben zu entnehmen war. Zur Fallbearbeitung benötigten die Teilneh-

---

<sup>1</sup> Siehe Anhang 1.

<sup>2</sup> Siehe Anhang 2.

<sup>3</sup> Dieses Verzeichnis findet sich auf der Homepage der Wirtschaftsprüferkammer und kann unter <http://www.wpk.de/wpverzeichnis/auswahl.asp> (Stand: 6.12.2006) abgerufen werden.

<sup>4</sup> Da aus Gründen der Vergleichbarkeit zu den bisherigen Teilnehmern lediglich Prüfer der Big Four-Prüfungsgesellschaften berücksichtigt wurden, sind die Anforderungen einer reinen Zufallsauswahl nicht erfüllt.

<sup>5</sup> Siehe Anhang 2.

<sup>6</sup> Um eine möglichst hohe Qualität des Falls zu gewährleisten, wurden die Fallunterlagen einem Partner sowie zwei Managern mit der Bitte um Überprüfung auf Konsistenz und Realitätsnähe überreicht. Die Anmerkungen wurden nach Rücksprache mit den jeweiligen Personen in die Fallunterlagen aufgenommen. Anschließend wurden vier Testläufe mit zwei Prüfungsassistenten und zwei Managern

mer ein weiteres Passwort, welches sie den heruntergeladenen Fallunterlagen entnehmen konnten.

Die Fallunterlagen<sup>1</sup> beinhalteten umfassende Hintergrundinformationen zu einem hypothetischen Unternehmen der Industriebranche, dessen Jahresabschlussdaten des zu prüfenden Geschäftsjahrs (ungeprüft) sowie des Vorjahrs (geprüft), darauf basierende jahresabschlussbasierte Kennzahlen und die Ergebnisse analytischer Prüfungshandlungen, die auf eine wesentliche unerwartete Abweichung in der Rohgewinnspanne des zu prüfenden Geschäftsjahrs hindeuteten. Die Teilnehmer wurden aufgefordert, ihre Aufmerksamkeit auf die Abweichung in der Rohgewinnspanne zu richten und den Grund für die festgestellte Abweichung zu identifizieren. Dabei wurden sie darauf hingewiesen, dass 90% der festgestellten Abweichung durch einen einzelnen Grund oder das mehrfache Auftreten des gleichen Grunds hervorgerufen wurde. Ferner wurde ihnen mitgeteilt, dass es sich bei dem Grund sowohl um einen Fehler im Rechnungswesen handeln kann als auch um Faktoren, die auf Änderungen in der Unternehmensumwelt zurückzuführen sind (sog. Nicht-Fehler).<sup>1</sup>

Die Hintergrundinformationen zum hypothetischen Mandanten beinhalteten u.a. Informationen über Unternehmensereignisse, die sich im zu prüfenden Geschäftsjahr ereignet haben. In den Fallversionen mit erhöhtem Prüfungsrisiko wurden die Teilnehmer darauf hingewiesen, dass im abgelaufenen Geschäftsjahr ein neuer Controller eingestellt wurde, da der alte Controller aus privaten Gründen das Unternehmen verlassen hatte. Die Fallunterlagen zu den Versionen mit niedrigem Prüfungsrisiko umfassten hingegen einen Hinweis, dass das Unternehmen im zu prüfenden Geschäftsjahr keine personellen oder strukturellen Änderungen erfahren hatte.

Bei der tatsächlichen Ursache der festgestellten Abweichung ist zwischen der Fehler- und der Nicht-Fehler-Version zu differenzieren. Im Rahmen der Fehler-Version wurde die Abweichung durch eine fehlerhafte Allokation der durch die gestiegenen Fertigungs- und Materialgemeinkosten verursachten Plankostenabweichung auf die Umsatzkosten und Vorratsbestände verursacht. Die entsprechende Plankostenabweichung wurde dabei vollständig auf die Vorratsbestände verbucht, anstatt sie verursachungsgerecht auf die Umsatzkosten und Vorratsbestände zu verteilen. Dieses führte zu einer Überbe-

---

durchgeführt. Alle Testläufe konnten erfolgreich abgeschlossen werden. Insbesondere ergaben sich keine Unklarheiten hinsichtlich der Fallbearbeitung und keine Verständnisprobleme.

<sup>1</sup> Siehe Anhang 4.

wertung der Vorräte und einer Unterbewertung der Umsatzkosten. Letztgenanntes führte zu einer Überbewertung der Rohgewinnspanne, welche das Verhältnis des Rohgewinns zu den Umsatzkosten misst. Im Rahmen der Nicht-Fehler-Version wurde die festgestellte Abweichung hingegen durch eine auf die gestiegene Nachfrage zurückzuführende Erhöhung der Absatzpreise verursacht, welche bei unterproportional gestiegenen Stückkosten zu einer Erhöhung der Rohgewinnspanne führte.

In den Fallversionen, in denen eine Mandantenbefragung simuliert wurde, umfassten die Fallunterlagen weiterhin eine Mandantenerklärung für die festgestellte Abweichung, welche die gestiegene Rohgewinnspanne auf die Erhöhung der Absatzpreise zurückführte. Im Rahmen der weiteren Fallbearbeitung konnten sowohl der Fehler- als auch der Nicht-Fehler-Version Hinweise entnommen werden, die eine vorgenommene Preiserhöhung vermuten ließen. Allerdings konnte die tatsächliche Abweichungsursache nur in der Nicht-Fehler-Version auf die Erhöhung der Absatzpreise zurückgeführt werden. Im Rahmen der Fehler-Version wurde die preissteigerungsbedingte Erhöhung des Rohgewinns hingegen durch die proportional gestiegenen Gemeinkosten kompensiert, so dass bei korrekten Jahresabschlussdaten keine Erhöhung der Rohgewinnspanne zu erwarten gewesen wäre.

Nachdem die Probanden die Fallunterlagen gelesen hatten, wurden sie dazu aufgefordert, sich mit dem zur Verfügung gestellten Passwort auf der entsprechenden Internetseite zu authentifizieren. Anschließend konnte die gesamte Fallbearbeitung online an einem beliebigen Personal Computer durchgeführt werden. Nach erfolgreicher Anmeldung sollten die Probanden zunächst potenzielle Abweichungshypothesen generieren und in eine Rangfolge von 1 bis N bringen<sup>2</sup>, wobei der Hypothese, die nach ihrer Einschätzung mit der höchsten Wahrscheinlichkeit für die festgestellte Abweichung verantwortlich ist, der Rang 1 zugeordnet werden sollte, und die Hypothese mit der geringsten Wahrscheinlichkeit mit Rang N gekennzeichnet werden sollte. Anschließend wurde den Probanden eine Liste mit 19 möglichen Prüfungstests bereitgestellt.<sup>1</sup> Die Probanden wählten zur Identifizierung der Abweichungsursache einen Test aus, nahmen die Testergebnisse zur Kenntnis und modifizierten – falls erwünscht – ihre vorangegangene Hypothesengenerierung einschließlich der Rangfolge. Diese Prozedur konnte von den Probanden solange wiederholt werden, bis sie der Ansicht waren, dass sie den

---

<sup>1</sup> Diese Vorgehensweise ist ebenfalls identisch zu den vorherigen Studien. Siehe etwa GREEN/TROTMAN (2003), S. 225 oder GREEN (2004b), S.10.

<sup>2</sup> N entspricht der Anzahl der generierten Hypothesen.



Grund für die festgestellte Abweichung identifiziert haben. Um die Prüfungspraxis möglichst realistisch nachzubilden, wurden die jeweiligen Prüfungstests mit einer virtuellen Prüfungsdauer gekennzeichnet und den Probanden ein Zeitbudget von 18 Stunden eingeräumt, welches nach Möglichkeit um nicht mehr als 10% überschritten werden sollte. Die Folgende Abbildung gibt den Ablauf der Fallbearbeitung graphisch wieder.

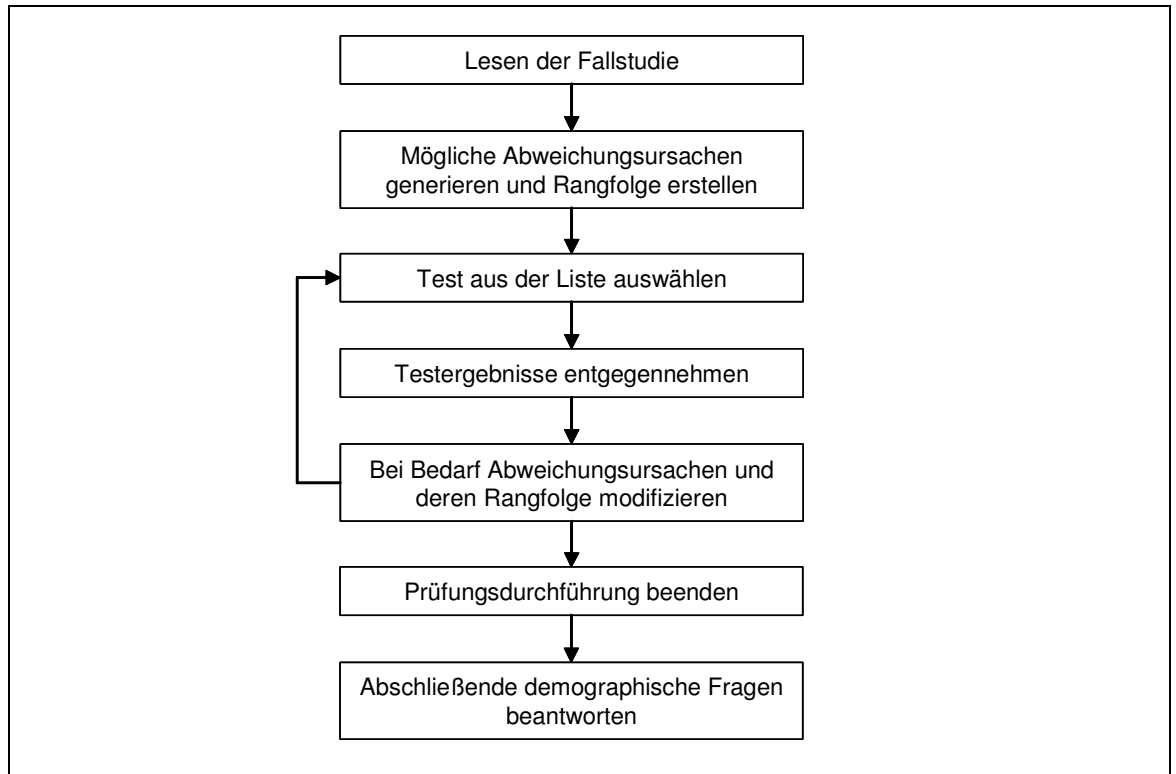


Abbildung 8: Graphische Darstellung der Vorgehensweise zur Fallbearbeitung

Die zur Verfügung gestellte Liste der 19 möglichen Tests sowie die entsprechenden Testergebnisse bezogen sich auf folgende fünf potenzielle Abweichungsursachen:

- Erhöhung der Absatzpreise;
- Fehlerhafte Allokation in der Plankostenabweichung;
- Fehler in den Umsatzerlösen;
- Allgemeine Änderungen in der internen oder externen Unternehmensumwelt (Nicht-Fehler);
- Fehler in den Beständen der Vorräte.

---

<sup>1</sup> Siehe Anhang 3.

Fünf der möglichen Tests richten sich direkt auf die Überprüfung der Erhöhung der Absatzpreise bzw. auf die Mandantenerklärung (Test 2, 8, 11, 12 und 18), vier auf die fehlerhafte Allokation der Plankostenabweichung (Test 5, 9, 13 und 17), drei auf einen Fehler in den Umsatzerlösen (Test 3, 7 und 14), zwei auf eine allgemeine Änderung (Test 1 und 19) und fünf auf einen Fehler in den Vorratsbeständen (Test 4, 6, 10, 15 und 16). Die Auswahl und die korrekte Interpretation der entsprechenden Tests sollte es den Probanden ermöglichen, den tatsächlichen Grund für die Abweichungsursache zu identifizieren.

## **4.4 Untersuchungsergebnisse**

### **4.4.1 Erläuterungen zu den durchgeführten Testverfahren**

Da die den in Kap. 4.4.2 bis 4.4.6 präsentierten Untersuchungsergebnissen zugrunde liegenden Daten nicht normalverteilt und somit die Voraussetzungen zur Durchführung von parametrischen Testverfahren nicht erfüllt sind, wurde auf die Durchführung von parametrischen Testverfahren zugunsten von nicht-parametrischen Testverfahren verzichtet.<sup>1</sup>

### **4.4.2 Manipulationscheck**

Um zu überprüfen, ob die fallinduzierte Variation des Prüfungsrisikos die gewünschte Wirkung auf die Probanden ausübte, wurden die Angaben zum empfundenen Prüfungsrisiko einer genaueren Analyse unterzogen.<sup>2</sup> Hierzu wurde die durchschnittliche Risikobeurteilung der Prüfergruppe, die im Rahmen ihrer Fallbearbeitung ein geringes Prüfungsrisiko zu berücksichtigen hatte, mit der durchschnittlichen Risikobeurteilung der Prüfergruppe verglichen, die einem erhöhten Prüfungsrisiko ausgesetzt war. Wie Tabelle 8 entnommen werden kann, schätzen Prüfer in der Gruppe mit erhöhtem Risiko das empfundene Prüfungsrisiko auf einem 5%-Signifikanzniveau wesentlich höher ein, als die Prüfergruppe mit geringem Prüfungsrisiko (3,13 Punkte ggü. 2,79 Punkte;  $p =$

---

<sup>1</sup> Siehe hierzu auch die Ausführungen in Kap. 3.4.3.

<sup>2</sup> Die Angaben zum empfundenen Prüfungsrisiko wurden nach der Fallbearbeitung im Rahmen des demographischen Fragenteils getätigt. Hierzu stand den Probanden eine 5 Punkte-Likert-Skala zur Verfügung, auf der sie das mit dem bearbeiteten Fall subjektiv wahrgenommene Prüfungsrisiko wiedergeben sollten (1 = niedriges Prüfungsrisiko; 5 = hohes Prüfungsrisiko).

0,019). Demnach kann davon ausgegangen werden, dass die Manipulation des Prüfungsrisikos erfolgreich war.

|                                                     | <b>Risikobeurteilung</b><br>$\mu$<br>[ $\sigma$ ] |
|-----------------------------------------------------|---------------------------------------------------|
| <b>Niedriges Prüfungsrisiko</b><br>n=42             | 2,79<br>[0,66]                                    |
| <b>Hohes Prüfungsrisiko</b><br>n=45                 | 3,13<br>[0,73]                                    |
| <b>Signifikanz<sup>1</sup></b><br><b>(1-seitig)</b> | 0,019                                             |

Tabelle 8: Risikobeurteilung der Prüfergruppe mit niedrigem und erhöhtem Prüfungsrisiko

### 4.4.3 Hypothesengenerierung

#### 4.4.3.1 Vorbemerkungen

Um einen möglichst hohen Informationsgehalt der Ergebnisse zu erhalten, wird im Rahmen der vorliegenden Studie zwischen der anfänglichen und der gesamten Hypothesengenerierung differenziert. Dabei ist zu beachten, dass bei der Untersuchung der anfänglichen Hypothesengenerierung nicht zwischen einer Fehler- und Nicht-Fehler-Version zu unterscheiden ist. Dieses ist darauf zurückzuführen, dass die Ausgangssituation in beiden Versionen identisch ist. Die Fallversionen unterscheiden sich erst im Hinblick auf die Testergebnisse, wobei im Rahmen der Fehler-Version die Tests zur Überprüfung der Allokation der Plankostenabweichung einen Fehler indizieren und die Tests bezogen auf die gestiegenen Absatzpreise lediglich scheinbar auf die Richtigkeit der „Mandantenerklärung“ hindeuten. Bei der Nicht-Fehler-Version zeigen die Tests zur Überprüfung der Plankostenallokation hingegen auf, dass keine fehlerhafte Verteilung vorliegt. Die Tests zur Überprüfung der Frage, ob die Abweichung auf die gestiegenen Absatzpreise zurückgeführt werden kann, indizieren hingegen die Gültigkeit dieser Annahme. Alle übrigen Testergebnisse sind bei beiden Fallversionen identisch. Aufgrund der unterschiedlichen Informationslage ab Beginn der Informationssuche<sup>2</sup> ist im Rahmen der Untersuchung der gesamten Hypothesengenerierung zwingend zwischen der

---

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Siehe hierzu Abbildung 1.

Fehler- und Nicht-Fehler-Version zu differenzieren, da davon auszugehen ist, dass die unterschiedlichen Testergebnisse die Generierung weiterer Hypothesen maßgeblich beeinflussen.

#### 4.4.3.2 Anfängliche Hypothesengenerierung

##### 4.4.3.2.1 Test von $H_{I,1}$ und $H_{I,2}$

Entgegen der Erwartungen von  $H_{I,1}$  und  $H_{I,2}$  lassen die Ergebnisse in Tabelle 9 weder einen signifikanten Einfluss des Erhalts einer Nicht-Fehlererklärung des Mandanten auf die Anzahl der generierten Fehlerhypothesen noch auf die Anzahl der generierten Nicht-Fehlerhypothesen erkennen. Prüfer mit Mandantenerklärung riefen im Schnitt 2,22 Fehlerhypothesen und 1,53 Nicht-Fehlerhypothesen ab im Vergleich zu 1,93 Fehler- und 1,33 Nicht-Fehlerhypothesen bei Prüfern ohne Mandantenerklärung. Da die Anzahl der generierten Fehlerhypothesen einerseits und Nicht-Fehlerhypothesen andererseits nicht signifikant voneinander abweichen ( $p = 0,32$  und  $p = 0,27$ ), sind  $H_{I,1}$  und  $H_{I,2}$  folglich abzulehnen. Demnach konnte in dieser Studie kein interference effect und spreading effect nachgewiesen werden.<sup>1</sup>

|                                                     | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|-----------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=44               | 2,22<br>[1,7]                             | 1,53<br>[1,37]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=43              | 1,93<br>[1,03]                            | 1,33<br>[1,25]                                  |
| <b>Signifikanz<sup>1</sup></b><br><b>(1-seitig)</b> | 0,32                                      | 0,27                                            |

Tabelle 9: Anfängliche Hypothesengenerierung der Prüfer

<sup>1</sup> Hingegen erreichte die Gruppe mit Mandantenerklärung auf einem 10%-Signifikanzniveau eine signifikant höhere Gesamtzahl an Abweichungshypothesen als Prüfer ohne Mandantenerklärung (3,75 ggü. 3,26;  $p = 0,09$ ; Mann-Whitney-U-Test; 2-seitige Signifikanz).

#### 4.4.3.2.2 Test von $H_{I.1.1}$ bis $H_{I.2.2}$

##### *Nicht-Spezialisten*

Wie Tabelle 10 zeigt, generierten Nicht-Spezialisten, die keine Mandantenerklärung erhielten, im Schnitt 2,25 Fehlerhypothesen und 1,30 Nicht-Fehlerhypothesen im Vergleich zu 1,95 bzw. 1,48 bei Nicht-Spezialisten mit Mandantenerklärung ( $p = 0,20$  bzw.  $p = 0,42$ ). Da der Einfluss des Erhalts einer Nicht-Fehlererklärung des Mandanten weder auf die Anzahl der generierten Fehlerhypothesen noch auf die Anzahl der generierten Nicht-Fehlerhypothesen signifikant ist, sind  $H_{I.1.1}$  und  $H_{I.2.1}$  abzulehnen. D.h., dass bei Nicht-Spezialisten im Rahmen dieser Studie weder der interference effect noch der spreading effect nachgewiesen werden konnte.<sup>2</sup>

|                                                | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung<br/>n=21</b>         | 1,95<br>[1,36]                            | 1,48<br>[1,57]                                  |
| <b>Ohne Mandantenerklärung<br/>n=20</b>        | 2,25<br>[0,97]                            | 1,30<br>[1,35]                                  |
| <b>Signifikanz<sup>3</sup><br/>(1- Seitig)</b> | 0,20                                      | 0,42                                            |

Tabelle 10: Anfängliche Hypothesengenerierung der Nicht-Spezialisten

##### *Spezialisten*

Wie Tabelle 11 entnommen werden kann, ist auch bei Spezialisten kein signifikanter Einfluss einer Mandantenerklärung auf die durchschnittliche Anzahl der generierten Nicht-Fehlerhypothesen zu verzeichnen. Spezialisten, denen eine Mandantenerklärung

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Darüber hinaus konnte bei den Nicht-Spezialisten ebenfalls kein Einfluss einer Mandantenerklärung auf die Anzahl der generierten Fehlerhypothesen einerseits und Nicht-Fehlerhypothesen andererseits festgestellt werden, wenn zwischen niedrigem und erhöhtem Prüfungsrisiko differenziert wurde. Bei niedrigem Prüfungsrisiko generierten Nicht-Spezialisten mit Mandantenerklärung (n=10) im Schnitt 2,1 ( $\sigma = 1,60$ ) Fehlerhypothesen und 1,60 ( $\sigma = 1,71$ ) Nicht-Fehlerhypothesen im Vergleich zu 2,3 ( $\sigma = 1,16$ ) und 1,00 ( $\sigma = 0,67$ ) bei Nicht-Spezialisten ohne Mandantenerklärung (n=10;  $p = 0,38$  bzw.  $p = 0,32$ ; Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz). Bei erhöhtem Prüfungsrisiko generierten Nicht-Spezialisten mit Mandantenerklärung (n=11) im Schnitt 1,82 ( $\sigma = 1,17$ ) Fehlerhypothesen und 1,36 ( $\sigma = 1,50$ ) Nicht-Fehlerhypothesen im Vergleich zu 2,2 ( $\sigma = 0,79$ ) und 1,60 ( $\sigma = 0,84$ ) bei Nicht-Spezialisten ohne Mandantenerklärung (n=10;  $p = 0,17$  bzw.  $p = 0,40$ ; Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz).

<sup>3</sup> Mann-Whitney-U-Test.

mitgeteilt wurde, generierten 1,57 Nicht-Fehlerhypothesen im Vergleich zu 1,35 bei Spezialisten ohne Mandantenerklärung ( $p = 0,28$ ). Somit ist  $H_{I.2.2}$  abzulehnen. Spezialisten konnte in dieser Studie kein spreading effect nachgewiesen werden.

|                                                      | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=23                | 2,48<br>[1,95]                            | 1,57<br>[1,20]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=23               | 1,65<br>[1,03]                            | 1,35<br>[1,15]                                  |
| <b>Signifikanz<sup>1</sup></b><br><b>(1- Seitig)</b> | 0,08                                      | 0,28                                            |

Tabelle 11: Anfängliche Hypothesengenerierung der Spezialisten

Unerwartet sind allerdings die Ergebnisse hinsichtlich der Anzahl der generierten Fehlerhypothesen. Zwar ist ein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf einem 10%-Signifikanzniveau zu erkennen. Allerdings gehen die Ergebnisse in eine andere als die in  $H_{I.1.2}$  vermutete Richtung. Spezialisten mit einer Nicht-Fehlererklärung des Mandanten generieren durchschnittlich signifikant mehr Fehlererklärungen als Spezialisten ohne Mandantenerklärung (2,48 ggü. 1,65;  $p = 0,08$ ).  $H_{I.1.2}$  ist somit abzulehnen. Vielmehr lassen die Ergebnisse bei Spezialisten einen „inversen“ spreading effect erkennen, wonach der Erhalt einer Nicht-Fehlererklärung des Mandanten zu einer vermehrten Generierung von Fehlerhypothesen führt. Ob die vermehrte Generierung von Fehlerhypothesen generell als positiv oder negativ für die weitere prüferische Vorgehensweise zu bezeichnen ist, hängt maßgeblich davon ab, ob es sich bei dem tatsächlichen Abweichungsgrund um einen Fehler oder Nicht-Fehler handelt. Da im Rahmen der anfänglichen Hypothesengenerierung nicht zwischen einer Fehler- und einer Nicht-Fehler-Version differenziert wird, kann an dieser Stelle keine weitere Bewertung vorgenommen werden.

Um die möglichen Ursachen dieser Beobachtung zu eruieren, wurden die Spezialisten in zwei Gruppen mit unterschiedlichem Prüfungsrisiko (hoch und niedrig) aufgeteilt. Dabei konnte festgestellt werden, dass der „inverse“ spreading effect ausschließlich bei der Gruppe mit erhöhtem Prüfungsrisiko zu verzeichnen war. Hier generierten die Spezia-

listen, denen eine Nicht-Fehlererklärung des Mandanten zur Verfügung gestellt wurde, auf einem 1%-Signifikanzniveau wesentlich mehr Fehlerhypothesen als Spezialisten ohne Mandantenerklärung (3,33 ggü. 1,50;  $p = 0,009^2$ ). Keinen Einfluss auf die Anzahl der generierten Fehlerhypothesen hatte die Mandantenerklärung hingegen bei der Gruppe mit niedrigem Prüfungsrisiko (1,55 ggü. 1,82;  $p = 0,28^3$ ).

Um das Verhalten der Spezialisten in der Gruppe mit erhöhtem Prüfungsrisiko zu begründen, wurde überprüft, ob diese bei Erhalt einer Nicht-Fehlererklärung des Mandanten eine unerwartete Reaktion erkennen ließen. Hierzu wurden die Angaben zum empfundenen Prüfungsrisiko<sup>4</sup> einer genaueren Untersuchung unterzogen. Die Analysen ließen erkennen, dass Spezialisten das Prüfungsrisiko auf einem 10%-Signifikanzniveau wesentlich höher einschätzen, wenn sie in Verbindung zum erhöhten Prüfungsrisiko eine Nicht-Fehlererklärung des Mandanten erhielten. Spezialisten, die sowohl ein erhöhtes Prüfungsrisiko als auch eine Mandantenerklärung zu berücksichtigen hatten, schätzten das Prüfungsrisiko auf einer 5-Punkte-Likert-Skala im Schnitt auf 3,38 Punkte ein im Vergleich zu einer Einschätzung von 3,04 Punkten bei der Spezialisten-Gruppe, die zwar einem erhöhten Prüfungsrisiko ausgesetzt war, zudem jedoch keine Nicht-Fehlererklärung des Mandanten erhielt ( $p = 0,10^5$ ). Demgegenüber konnte bei den Spezialisten kein signifikanter Einfluss einer Nicht-Fehlererklärung des Mandanten auf das empfundene Prüfungsrisiko festgestellt werden, wenn das zu berücksichtigende Prüfungsrisiko gering ausfiel. (2,59 ggü. 2,79;  $p = 0,33^6$ ). Diese Ergebnisse lassen darauf schließen, dass Spezialisten in einer Prüfungssituation mit erhöhtem Prüfungsrisiko besonders sensibel auf die Mitteilung einer Nicht-Fehlererklärung reagieren. Es ist zu vermuten, dass diese Reaktion auf die erhöhte professionelle Skepsis von Spezialisten in der Industriebranche zurückzuführen ist.

---

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>3</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>4</sup> Diese Angaben wurden nach der Fallbearbeitung im Rahmen des demographischen Fragenteils getätigt. Hierzu stand den Probanden eine 1 – 5 Likert-Skala zu Verfügung, auf der sie das mit dem bearbeiteten Fall verbundene Prüfungsrisiko kennzeichnen sollten (1 = niedriges Prüfungsrisiko; 5 = hohes Prüfungsrisiko).

<sup>5</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>6</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

#### 4.4.3.2.3 Test von $FF_{I,1}$

Tabelle 12 zeigt, dass Prüfer mit Mandantenerklärung bei geringem Prüfungsrisiko auf einem 10%-Signifikanzniveau wesentlich mehr Nicht-Fehlerhypothesen generierten als Probanden ohne Mandantenerklärung (1,76 ggü. 1,14;  $p = 0,10^1$ ). Bei erhöhtem Prüfungsrisiko hingegen ist der vermehrte Abruf von Nicht-Fehlerhypothesen nicht evident (1,30 ggü. 1,50,  $p = 0,37^2$ ). Demnach weisen die Ergebnisse darauf hin, dass die anfängliche Hypothesengenerierung lediglich unter geringem Prüfungsrisiko vom spreading effect betroffen zu sein scheint.<sup>3</sup> Erhöhtes Prüfungsrisiko scheint demnach einen geeigneten debiasing-Mechanismus gegen den spreading effect darzustellen.

|                            |                 | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|----------------------------|-----------------|-------------------------------------------|-------------------------------------------------|
| Mit<br>Mandantenerklärung  | Hoch<br>n=23    | 2,61<br>[1,90]                            | 1,30<br>[1,26]                                  |
|                            | Niedrig<br>n=21 | 1,81<br>[1,36]                            | 1,76<br>[1,48]                                  |
| Ohne<br>Mandantenerklärung | Hoch<br>n=22    | 1,82<br>[0,91]                            | 1,50<br>[1,47]                                  |
|                            | Niedrig<br>n=21 | 2,05<br>[1,16]                            | 1,14<br>[0,96]                                  |

Tabelle 12: Anfängliche Hypothesengenerierung der Prüfer unter Berücksichtigung des Prüfungsrisikos

Darüber hinaus ist zu erkennen, dass auch die anfängliche Generierung von Fehlerhypothesen unterschiedliche Ergebnisse hervorbringt, wenn zwischen niedrigem und erhöhtem Prüfungsrisiko differenziert wird. Bei erhöhtem Prüfungsrisiko generieren Prüfer mit Mandantenerklärung auf einem 10%-Signifikanzniveau wesentlich mehr Fehlerhypothesen als Prüfer ohne Mandantenerklärung (2,61 ggü. 1,82;  $p = 0,09^4$ ).<sup>1</sup> Hingegen generierten Prüfer mit Mandantenerklärung nicht signifikant mehr Fehler-Hypothesen

<sup>1</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>2</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>3</sup> Eine detaillierte Beschreibung des spreading effects erfolgt in Kap. 3.5.3.3.1.2.1 i.V.m. 3.5.3.3.2.1.2.2.

<sup>4</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.



als Prüfer ohne Mandantenerklärung, wenn lediglich ein geringes Prüfungsrisiko zu berücksichtigen war (1,81 ggü. 2,05;  $p = 0,26^2$ ).

#### 4.4.3.3 Gesamte Hypothesengenerierung

##### 4.4.3.3.1 Test von $H_{I,3}$ bis $H_{I,4}$

##### 4.4.3.3.1.1 Fehler-Version

Wie Tabelle 13 zu entnehmen ist, generierten Prüfer mit Mandantenerklärung im Schnitt 2,68 Fehlerhypothesen und 1,55 Nicht-Fehlerhypothesen. Prüfer ohne Mandantenerklärung riefen demgegenüber durchschnittlich 2,29 Fehler-Hypothesen und 1,86 Nicht-Fehlerhypothesen ab. Die Unterschiede zwischen den beiden Prüfergruppen waren nicht signifikant ( $p = 0,38$  und  $p = 0,30$ ).<sup>3</sup> Demnach sind  $H_{I,3}$  und  $H_{I,4}$  im Rahmen der Fehler-Version abzulehnen.

|                                                     | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|-----------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=22               | 2,68<br>[1,84]                            | 1,55<br>[1,41]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=21              | 2,29<br>[1,10]                            | 1,86<br>[1,65]                                  |
| <b>Signifikanz<sup>4</sup></b><br><b>(1-seitig)</b> | 0,38                                      | 0,30                                            |

Tabelle 13: Gesamte Hypothesengenerierung der Prüfer (Fehler-Version)

##### 4.4.3.3.1.2 Nicht-Fehler-Version

Tabelle 14 zeigt, dass Prüfer mit Mandantenerklärung durchschnittlich 2,45 Fehler- und 2,14 Nicht-Fehlerhypothesen abriefen. Im Vergleich hierzu generierten Prüfer ohne Mandantenerklärung im Schnitt 2,36 Fehler- und 1,41 Nicht-Fehlerhypothesen. Dem-

<sup>1</sup> Eine genauere Untersuchung dieses Phänomens erfolgte bereits in Kap. 4.4.3.2.2, da diese Beobachtung ausschließlich auf das Verhalten der Spezialisten zurückzuführen ist.

<sup>2</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>3</sup> Zudem ergab sich kein signifikanter Einfluss auf die Gesamtzahl der generierten Hypothesen (4,23 ggü. 4,14;  $p = 0,84$ ; Mann-Whitney-U-Test; 2-seitige Signifikanz).

<sup>4</sup> Mann-Whitney-U-Test.

nach generierten Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhielten, auf einem 5%-Signifikanzniveau wesentlich mehr Nicht-Fehlerhypothesen als Prüfer ohne Mandantenerklärung ( $p = 0,046$ ). Somit ist  $H_{I,4}$  im Rahmen der Nicht-Fehler-Version anzunehmen. Die gesamte Hypothesengenerierung scheint vom spreading effect gekennzeichnet zu sein. Keine signifikanten Unterschiede konnten hingegen bei der Anzahl der abgerufenen Fehlerhypothesen festgestellt werden ( $p = 0,39$ ). Folglich ist  $H_{I,3}$  abzulehnen.<sup>1</sup>

|                                        | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|----------------------------------------|-------------------------------------------|-------------------------------------------------|
| Mit Mandantenerklärung<br>n=22         | 2,45<br>[1,68]                            | 2,14<br>[1,42]                                  |
| Ohne Mandantenerklärung<br>n=22        | 2,36<br>[1,36]                            | 1,41<br>[1,05]                                  |
| Signifikanz <sup>1</sup><br>(1-seitig) | 0,39                                      | 0,046                                           |

Tabelle 14: Gesamte Hypothesengenerierung der Prüfer (Nicht-Fehler-Version)

#### 4.4.3.3.2 Test von $H_{I,3,1}$ bis $H_{I,4,2}$

##### 4.4.3.3.2.1 Fehler-Version

###### *Nicht-Spezialisten*

Entgegen der Erwartungen von  $H_{I,3,1}$  und  $H_{I,4,1}$  weisen die Ergebnisse in Tabelle 15 darauf hin, dass die gesamte Hypothesengenerierung der Nicht-Spezialisten keinem signifikanten Einfluss der Mandantenerklärung unterliegt. Weder die durchschnittliche Anzahl der Fehlerhypothesen noch der Nicht-Fehlerhypothesen weichen signifikant voneinander ab (2,44 ggü. 2,33;  $p = 0,46/1,67$  ggü. 1,56;  $p = 0,43$ ). Demnach ist zu konstatieren, dass  $H_{I,3,1}$  und  $H_{I,4,1}$  abzulehnen sind.

---

<sup>1</sup> Es konnte zudem festgestellt werden, dass der Erhalt einer Mandantenerklärung auf einem 10%-Signifikanzniveau einen wesentlichen Einfluss auf die Gesamtzahl der generierten Abweichungshypo-

|                                                      | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=9                 | 2,44<br>[1,42]                            | 1,67<br>[1,80]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=9                | 2,33<br>[1,22]                            | 1,56<br>[1,81]                                  |
| <b>Signifikanz<sup>2</sup></b><br><b>(1- Seitig)</b> | 0,46                                      | 0,43                                            |

Tabelle 15: Gesamte Hypothesengenerierung der Nicht-Spezialisten (Fehler-Version)

### *Spezialisten*

Vergleichbar zu den Ergebnissen der Nicht-Spezialisten lassen auch die Ergebnisse der Spezialisten keinen signifikanten Einfluss des Erhalts einer Mandantenerklärung auf die durchschnittliche Anzahl der Fehlerhypothesen einerseits und Nicht-Fehlerhypothesen andererseits erkennen. Spezialisten mit Mandantenerklärung riefen im Rahmen der Fehler-Version durchschnittlich 2,85 Fehlerhypothesen und 1,46 Nicht-Fehlerhypothesen ab. Demgegenüber generierten Spezialisten ohne Mandantenerklärung im Schnitt 2,25 Fehlerhypothesen und 2,08 Nicht-Fehlerhypothesen. Da die Gruppenunterschiede jeweils nicht signifikant sind ( $p = 0,41$  bzw.  $p = 0,16$ ), sind  $H_{I,3,2}$  sowie  $H_{I,4,2}$  abzulehnen.

|                                                      | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=13                | 2,85<br>[2,12]                            | 1,46<br>[1,13]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=12               | 2,25<br>[1,06]                            | 2,08<br>[1,56]                                  |
| <b>Signifikanz<sup>3</sup></b><br><b>(1- Seitig)</b> | 0,41                                      | 0,16                                            |

Tabelle 16: Gesamte Hypothesengenerierung der Spezialisten (Fehler-Version)

thesen ausübte ( $p = 0,066$ ; Mann-Whitney-U-Test; 2-seitige Signifikanz).

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Mann-Whitney-U-Test.

<sup>3</sup> Mann-Whitney-U-Test.

Darüber hinaus ist darauf hinzuweisen, dass der im Rahmen der anfänglichen Hypothesengenerierung zu verzeichnende „inverse“ spreading effect<sup>1</sup> bei der gesamten Hypothesengenerierung nicht mehr nachgewiesen werden kann. Die Ergebnisse gehen zwar weiterhin in die eines „inversen“ spreading effects entsprechende Richtung; Spezialisten mit Nicht-Fehlererklärung des Mandanten generieren im Schnitt 2,85 Fehlerhypothesen im Vergleich zu 2,25 Fehlerhypothesen bei Spezialisten ohne Mandantenerklärung. Allerdings sind diese Unterschiede nicht (mehr) signifikant ( $p = 0,41$ ). Folglich könnte die Vermutung aufgestellt werden, dass die durch den Erhalt einer Mandantenerklärung vermehrte anfängliche Generierung von Fehlerhypothesen durch die verringerte nachfolgende Generierung von Fehlerhypothesen kompensiert wird.

#### 4.4.3.3.2.2 Nicht-Fehler-Version

##### *Nicht-Spezialisten*

Wie in Tabelle 17 zu erkennen ist, weisen die Ergebnisse darauf hin, dass der Erhalt einer Mandantenerklärung weder einen signifikanten Einfluss auf die Anzahl der abgerufenen Fehlerhypothesen noch auf die Anzahl der generierten Nicht-Fehlerhypothesen ausübt. Zwar gehen die Ergebnisse in die durch  $H_{I.3.1}$  und  $H_{I.4.1}$  unterstellte Richtung. Dennoch sind  $H_{I.3.1}$  und  $H_{I.4.1}$  auf einem 10%-Signifikanzniveau abzulehnen. Analog zur Nicht-Fehler-Versionen lassen die Befunde nicht erkennen, dass die gesamte Hypothesengenerierung der Nicht-Spezialisten durch einen interference effect oder durch einen spreading effect geprägt ist.

|                                                      | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------------|-------------------------------------------|-------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=12                | 2,50<br>[1,73]                            | 1,92<br>[1,44]                                  |
| <b>Ohne Mandantenerklärung</b><br>n=11               | 3,00<br>[1,34]                            | 1,18<br>[0,98]                                  |
| <b>Signifikanz<sup>1</sup></b><br><b>(1- Seitig)</b> | 0,24                                      | 0,11                                            |

Tabelle 17: Gesamte Hypothesengenerierung der Nicht-Spezialisten (Nicht-Fehler-Version)

<sup>1</sup> Vgl. hierzu Kap. 4.4.3.2.2.

## Spezialisten

Entgegen der Erwartung von  $H_{I.3.2}$  zeigt Tabelle 18, dass Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, nicht signifikant weniger Fehlerhypothesen generieren als Spezialisten ohne Mandantenerklärung (2,40 ggü. 1,73;  $p = 0,15$ ). Vielmehr kann festgestellt werden, dass der im Rahmen der anfänglichen Hypothesengenerierung nachgewiesene „inverse“ spreading effect zwar nicht mehr signifikant, jedoch in die entsprechende Richtung zu verzeichnen ist ( $p = 0,15$ ).  $H_{I.3.2}$  ist demnach auch im Rahmen der gesamten Hypothesengenerierung abzulehnen.

|                                                      | <b>Fehlerhypothesen</b><br>$\mu$<br>[ $\sigma$ ] | <b>Nicht-Fehlerhypothesen</b><br>$\mu$<br>[ $\sigma$ ] |
|------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------|
| <b>Mit Mandantenerklärung</b><br>n=10                | 2,40<br>[1,71]                                   | 2,40<br>[1,43]                                         |
| <b>Ohne Mandantenerklärung</b><br>n=11               | 1,73<br>[1,10]                                   | 1,64<br>[1,12]                                         |
| <b>Signifikanz<sup>2</sup></b><br><b>(1- Seitig)</b> | 0,15                                             | 0,10                                                   |

Tabelle 18: Gesamte Hypothesengenerierung der Spezialisten (Nicht-Fehler-Version)

In Bezug auf die Anzahl der Nicht-Fehlerhypothesen ist allerdings festzuhalten, dass die Ergebnisse auf einem 10%-Signifikanzniveau einen wesentlichen Einfluss des Erhalts einer Mandantenerklärung erkennen lassen. Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhielten, riefen im Schnitt signifikant mehr Nicht-Fehlerhypothesen als Spezialisten ohne Mandantenerklärung ab ( $p = 0,10$ ). Folglich ist im Rahmen dieser Studie die gesamte Hypothesengenerierung der Spezialisten in der Nicht-Fehler-Version durch den spreading effect gekennzeichnet.  $H_{I.4.2}$  ist demnach anzunehmen. Da es sich in der betrachteten Situation jedoch um eine Nicht-Fehler-Version handelt, kann nicht konstatiert werden, dass hierdurch ein negativer Einfluss auf die weitere prüferische Vorgehensweise zu erwarten ist.

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Mann-Whitney-U-Test.

#### 4.4.3.3.3 Test von $FF_{1,2}$

##### 4.4.3.3.3.1 Fehler-Version

Wie aus Tabelle 19 zu entnehmen ist, konnten weder bei niedrigem noch bei erhöhtem Prüfungsrisiko signifikante Unterschiede zwischen Prüfern mit und Prüfern ohne Mandantenerklärung festgestellt werden. Probanden mit Mandantenerklärung riefen weder bei geringem noch bei erhöhtem Prüfungsrisiko weniger Fehlerhypothesen (2,36 ggü. 2,36;  $p = 0,41$  bzw. 3,00 ggü. 2,20;  $p = 0,27$ ) bzw. mehr Nicht-Fehlerhypothesen (1,64 ggü. 1,64;  $p = 0,43$  bzw. 1,45 ggü. 2,10;  $p = 0,19$ ) als Probanden ohne Mandantenerklärung ab.<sup>1</sup> Demnach konnte im Rahmen der Fehler-Version kein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss einer Mandantenerklärung auf die gesamte Hypothesengenerierung nachgewiesen werden.

|                            |                 | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|----------------------------|-----------------|-------------------------------------------|-------------------------------------------------|
| Mit<br>Mandantenerklärung  | Hoch<br>n=11    | 3,00<br>[2,28]                            | 1,45<br>[1,44]                                  |
|                            | Niedrig<br>n=11 | 2,36<br>[1,29]                            | 1,64<br>[1,43]                                  |
| Ohne<br>Mandantenerklärung | Hoch<br>n=10    | 2,20<br>[0,92]                            | 2,10<br>[1,73]                                  |
|                            | Niedrig<br>n=11 | 2,36<br>[1,29]                            | 1,64<br>[1,63]                                  |

Tabelle 19: Gesamte Hypothesengenerierung der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

##### 4.4.3.3.3.2 Nicht-Fehler-Version

Tabelle 20 zeigt, dass die Prüfer sowohl bei niedrigem als auch bei erhöhtem Prüfungsrisiko auf einem 10%-Signifikanzniveau wesentlich mehr Nicht-Fehlerhypothesen generieren, wenn ihnen eine Mandantenerklärung zur Verfügung gestellt wird (2,40 ggü. 1,70;  $p = 0,10$  bzw. 1,17 ggü. 1,92;  $p = 0,09$ ).<sup>1</sup> Hingegen ist weder bei niedrigem noch bei erhöhtem Prüfungsrisiko ein signifikanter Einfluss des Erhalts einer Mandantener-

<sup>1</sup> Jeweils Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

klärung auf die Anzahl der abgerufenen Fehlerhypothesen zu verzeichnen (1,90 ggü. 2,20;  $p = 0,31$  bzw. 2,92 ggü. 2,50; 0,22).<sup>2</sup> Folglich ist festzuhalten, dass im Rahmen der Nicht-Fehler-Version kein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss einer Mandantenerklärung auf die gesamte Hypothesengenerierung nachgewiesen werden konnte.

|                            |                 | Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] | Nicht-Fehlerhypothesen<br>$\mu$<br>[ $\sigma$ ] |
|----------------------------|-----------------|-------------------------------------------|-------------------------------------------------|
| Mit<br>Mandantenerklärung  | Hoch<br>n=12    | 2,92<br>[1,51]                            | 1,92<br>[1,44]                                  |
|                            | Niedrig<br>n=10 | 1,90<br>[1,79]                            | 2,40<br>[1,43]                                  |
| Ohne<br>Mandantenerklärung | Hoch<br>n=12    | 2,50<br>[1,45]                            | 1,17<br>[1,27]                                  |
|                            | Niedrig<br>n=10 | 2,20<br>[1,32]                            | 1,70<br>[0,67]                                  |

Tabelle 20: Gesamte Hypothesengenerierung der Prüfer mit Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version)

#### 4.4.4 Informationssuche

##### 4.4.4.1 Vorbemerkungen

Zur Überprüfung von  $H_{II.1}$  sowie  $H_{II.1.1}$  und  $H_{II.1.2}$  wird auf die Anzahl der Tests zurückgegriffen, die sich auf die Überprüfung der Mandantenerklärung beziehen (Kat. 1-Tests). Zur Überprüfung von  $H_{II.2}$  sowie  $H_{II.2.1}$  und  $H_{II.2.2}$  soll hingegen auf die Anzahl der geprüften Abweichungshypothesen (Breite) zurückgegriffen werden.

<sup>1</sup> Jeweils Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>2</sup> Jeweils Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

#### 4.4.4.2 Der Einfluss auf die angewandte Suchstrategie

##### 4.4.4.2.1 Test von $H_{II,1}$ und $H_{II,2}$

###### 4.4.4.2.1.1 Fehler-Version

Entgegen der Vermutung von  $H_{II,1}$  konnten in Abhängigkeit des Erhalts einer Mandantenerklärung keine signifikanten Unterschiede zwischen der jeweiligen Anzahl der durchgeführten Kat. 1-Tests festgestellt werden (siehe Tabelle 21). Prüfer mit Mandantenerklärung führten im Schnitt 1,64 und Prüfer ohne Mandantenerklärung 1,38 Kat. 1-Tests durch ( $p = 0,21$ ). Obwohl die Ergebnisse somit in die erwartete Richtung weisen, ist  $H_{II,1}$  abzulehnen. Ferner ist Tabelle 21 zu entnehmen, dass Prüfer, denen eine Mandantenerklärung zur Verfügung gestellt wurde, durchschnittlich 3,32 potenzielle Abweichungsursachen untersuchten. Probanden ohne Mandantenerklärung gingen demgegenüber im Schnitt 3,10 möglichen Abweichungsursachen nach. Da sich die Breite der angewandten Suchstrategie der Prüfer mit und ohne Mandantenerklärung somit nicht signifikant voneinander unterscheidet, ist folglich auch  $H_{II,2}$  abzulehnen ( $p = 0,57$ ).<sup>1</sup>

Bezüglich der Anzahl der durchgeführten Tests pro möglicher Abweichungsursache (Tiefe) ließ sich ein signifikanter Unterschied zwischen Prüfern mit und ohne Mandantenerklärung lediglich hinsichtlich Kat. 4 feststellen (0,05 ggü. 0,19;  $p = 0,07^2$ ).<sup>3</sup> Da es sich hierbei um Tests zur Überprüfung eines Nicht-Fehlers handelt, wäre aufgrund theoretischer Überlegungen<sup>4</sup> zu vermuten, dass Probanden, die eine Nicht-Fehlererklärung des Mandanten erhalten, diese Tests wesentlich häufiger durchführen. Hier jedoch werden diese Tests häufiger von Probanden durchgeführt, die keine Mandantenerklärung erhalten.<sup>5</sup>

---

<sup>1</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>2</sup> Mann-Whitney-U-Test (1-seitige Signifikanz)

<sup>3</sup> Die Signifikanzen der übrigen Kategorien betragen (Mann-Whitney-U-Test; 1-seitige Signifikanz): Kat. 1 ( $p = 0,21$ ); Kat. 3 ( $p = 0,13$ ); Kat. 5 ( $p = 0,21$ ).

<sup>4</sup> Da im Rahmen der vorliegenden Studie davon ausgegangen wird, dass der Erhalt einer Nicht-Fehlererklärung zu einer vermehrten Generierung von weiteren Nicht-Fehlerhypothesen führt (spreading effect), wäre anzunehmen, dass dieses im Rahmen der Informationssuche zu einer verstärkten Durchführung von Nicht-Fehlerhypothesen untersuchenden Tests führt.

<sup>5</sup> Eine genauere Untersuchung dieser Beobachtung erfolgt in Kap. 4.4.4.2.3.1, da diese Beobachtung ausschließlich auf das Verhalten der Nicht-Spezialisten zurückzuführen ist.



|                                                         | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=21 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| Breite<br>$\mu$<br>[ $\sigma$ ]                         | 3,32<br>[0,84]                    | 3,10<br>[1,22]                     | 0,57                                   |
| Kat. 1<br>(Mandantenerklärung)<br>$\mu$<br>[ $\sigma$ ] | 1,64<br>[0,95]                    | 1,38<br>[1,07]                     | 0,21                                   |

Tabelle 21: Angewandte Suchstrategie der Prüfer (Fehler-Version)

#### 4.4.4.2.1.2 Nicht-Fehler-Version

Vergleichbar zur Fehler-Version können  $H_{II.1}$  und  $H_{II.2}$  auch in der Nicht-Fehler-Version nicht bestätigt werden. Wie Tabelle 22 zu entnehmen ist, führten Prüfer, die eine Mandantenerklärung erhielten, durchschnittlich 1,77 Kat. 1-Tests durch und deckten durchschnittlich 3,32 Abweichungshypothesen ab. Probanden ohne Mandantenerklärung wählten demgegenüber 1,50 Kat. 1-Tests aus und untersuchten durchschnittlich 3,09 Abweichungsursachen. Demnach ist festzuhalten, dass der Erhalt einer Mandantenerklärung weder einen signifikanten Einfluss auf die Anzahl der durchgeführten Kat. 1-Tests ( $p = 0,16$ ) noch auf die Anzahl der untersuchten Abweichungshypothesen (Breite) ausübt ( $p = 0,37$ ).<sup>2</sup> Folglich sind  $H_{II.1}$  und  $H_{II.2}$  auch im Rahmen der Nicht-Fehler-Version abzulehnen.<sup>3</sup>

Bezüglich der Anzahl der durchgeführten Tests pro Kategorie konnte ein signifikanter Einfluss der Mandantenerklärung lediglich hinsichtlich Kat. 5 festgestellt werden. Hier führten Probanden mit Mandantenerklärung auf einem 1%-Signifikanzniveau wesentlich mehr Tests durch als Probanden ohne Mandantenerklärung (1,41 ggü. 0,68;  $p = 0,002^4$ ).<sup>5</sup>

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Jeweils Mann-Whitney-U-Test (1-seitige Signifikanz).

<sup>3</sup> Allerdings konnte festgestellt werden, dass Probanden mit Mandantenerklärung auf einem 10%-Signifikanzniveau wesentlich mehr Tests durchführten (5,82 ggü. 5,00;  $p = 0,10$ ) und dabei auf einem 5%-Signifikanzniveau wesentlich mehr Zeit in Anspruch nahmen (16,73 Std. ggü. 13,95 Std.;  $p = 0,03$ ). Jeweils Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>4</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>5</sup> Für eine genauere Analyse dieser Beobachtung soll an dieser Stelle auf Kap. 4.4.4.2.3.2 verwiesen werden, da bei Spezialisten einerseits und Nicht-Spezialisten andererseits scheinbar unterschiedliche Ursachen zu diesem Effekt geführt haben.

|                                                         | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=22 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| Breite<br>$\mu$<br>[ $\sigma$ ]                         | 3,32<br>[0,84]                    | 3,09<br>[1,06]                     | 0,37                                   |
| Kat. 1<br>(Mandantenerklärung)<br>$\mu$<br>[ $\sigma$ ] | 1,77<br>[0,92]                    | 1,50<br>[0,96]                     | 0,16                                   |

Tabelle 22: Angewandte Suchstrategie der Prüfer (Nicht-Fehler-Version)

#### 4.4.4.2.2

#### 4.4.4.2.3 Test von $H_{II.1.1}$ bis $H_{II.2.2}$

##### 4.4.4.2.3.1 Fehler-Version

##### *Nicht-Spezialisten*

Entgegen der Erwartung von  $H_{II.1.1}$  führt der Erhalt einer Nicht-Fehlererklärung des Mandanten nicht zu einer verstärkten Einholung von Informationen, die der Beurteilung der Mandantenerklärung dienen (siehe Tabelle 23). Nicht-Spezialisten mit Mandantenerklärung führten durchschnittlich 1,01 Kat. 1-Tests durch im Vergleich zu 1,56 Kat. 1-Tests bei Nicht-Spezialisten ohne Mandantenerklärung. Demzufolge ist  $H_{II.1.1}$  abzulehnen ( $p = 0,42$ ). Darüber hinaus weisen die Ergebnisse ebenfalls darauf hin, dass Nicht-Spezialisten mit Mandantenerklärung nicht signifikant weniger Hypothesen testeten als Nicht-Spezialisten ohne Mandantenerklärung (3,44 ggü. 3,22;  $p = 0,48$ ). Somit ist auch  $H_{II.2.1}$  abzulehnen. Ein Einfluss der Mandantenerklärung auf die Eigenschaften der Informationssuche konnte bei Nicht-Spezialisten in der Fehler-Version nicht verzeichnet werden.

---

<sup>1</sup> Mann-Whitney-U-Test.

|                                                                       | Mit<br>Mandantenerklärung<br>n=9 | Ohne<br>Mandantenerklärung<br>n=9 | Signifikanz <sup>1</sup><br>(1-seitig) |
|-----------------------------------------------------------------------|----------------------------------|-----------------------------------|----------------------------------------|
| <b>Breite</b><br>$\mu$<br>[ $\sigma$ ]                                | 3,44<br>[0,53]                   | 3,22<br>[1,09]                    | 0,48                                   |
| <b>Kat. 1</b><br><b>(Mandantenerklärung)</b><br>$\mu$<br>[ $\sigma$ ] | 1,01<br>[0,92]                   | 1,56<br>[1,13]                    | 0,42                                   |

Tabelle 23: Breite der Suche und Tiefe der Kat. 1-Tests der Nicht-Spezialisten (Fehler-Version)

Allerdings zeigt Tabelle 24, dass Nicht-Spezialisten, denen eine Nicht-Fehlererklärung des Mandanten zur Verfügung gestellt wurde, auf einem 10%-Signifikanzniveau wesentlich mehr Kat. 5- bzw. Fehler-Tests (0,89 ggü. 0,56;  $p = 0,06$ ) und wesentlich weniger Kat. 4- bzw. Nicht-Fehler-Tests (0,00 ggü. 0,22;  $p = 0,07$ ) durchführten.<sup>2</sup> Dieses ist jedoch überraschend, da in der vorliegenden Studie davon ausgegangen wird, dass die Vererbung einer Nicht-Fehlererklärung zu einer verstärkten Generierung von Nicht-Fehlerhypothesen und einem eingeschränkten Abruf von Fehlerhypothesen führt und dieses im Rahmen der Informationssuche eine vermehrte (verringerte) Einholung von Prüfungsnachweisen zu Nicht-Fehlerhypothesen (Fehlerhypothesen) bewirkt. Die vorliegenden Ergebnisse weisen jedoch in die entgegengesetzte Richtung.

|                                                                           | Mit<br>Mandantenerklärung<br>n=9 | Ohne<br>Mandantenerklärung<br>n=9 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------------------------|----------------------------------|-----------------------------------|----------------------------------------|
| <b>Kat. 4</b><br><b>(Nicht-Fehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,00<br>[0,00]                   | 0,22<br>[0,44]                    | 0,07                                   |
| <b>Kat. 5</b><br><b>(Fehler in den Vorräten)</b><br>$\mu$<br>[ $\sigma$ ] | 0,89<br>[0,33]                   | 0,56<br>[0,53]                    | 0,06                                   |

Tabelle 24: Tiefe der Kat. 4 & 5-Tests der Nicht-Spezialisten (Fehler-Version)

Um eine Begründung für das Verhalten der Nicht-Spezialisten zu finden, wurde untersucht, ob der Erhalt einer Mandantenerklärung zu einer unerwarteten Auswirkung führ-

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Hinsichtlich des Suchumfangs sowie der Suchtiefe in Bezug auf Kat. 2 und Kat. 3 konnten keine signifikanten Unterschiede festgestellt werden.

te. Hierzu wurden die Angaben zum empfundenen Prüfungsrisiko der Nicht-Spezialisten in Abhängigkeit des Erhalts einer Mandantenerklärung einer genaueren Betrachtung unterzogen. Dabei fiel auf, dass der Erhalt einer Mandantenerklärung ein erhöhtes wahrgenommenes Prüfungsrisiko mit sich brachte (siehe Tabelle 25). Nicht-Spezialisten beurteilten das Prüfungsrisiko auf einer 5 Punkte-Likert-Skala mit durchschnittlich 3,03 Punkten im Vergleich zu 2,87 Punkten bei Nicht-Spezialisten ohne Mandantenerklärung. Möglicherweise führte die professionelle Skepsis der Nicht-Spezialisten zu einer erhöhten Risikowahrnehmung, wenn ihnen eine Nicht-Fehlererklärung des Mandanten zur Verfügung gestellt wurde. Obwohl der Einfluss nicht signifikant ist ( $p = 0,12$ ), könnte dieses dazu geführt haben, dass der Erhalt einer Mandantenerklärung – und hiermit verbunden eine erhöhte Risikowahrnehmung – zu einer vermehrten Durchführung von Kat. 5-Tests (Fehler) und einer verringerten Durchführung von Kat. 4-Tests (Nicht-Fehler) geführt hat.

|                                                           | Mit<br>Mandantenerklärung<br>n=21 | Ohne<br>Mandantenerklärung<br>n=20 | Signifikanz <sup>2</sup><br>(1-seitig) |
|-----------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| Wahrgenommenes<br>Prüfungsrisiko<br>$\mu$<br>[ $\sigma$ ] | 3,03<br>[0,72]                    | 2,87<br>[0,59]                     | 0,12                                   |

Tabelle 25: Empfundenes Prüfungsrisiko der Nicht-Spezialisten in Abhängigkeit des Erhalts einer Mandantenerklärung

### *Spezialisten*

Entsprechend  $H_{II.1.2}$  weisen die Ergebnisse in Tabelle 26 darauf hin, dass Spezialisten auf einem 10%-Signifikanzniveau wesentlich mehr Tests zur Einholung von Informationen durchführten, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Spezialisten ohne Mandantenerklärung (1,77 ggü. 1,25;  $p = 0,09$ ). Somit ist  $H_{II.1.2}$  zu bestätigen. Keinen signifikanten Einfluss hatte der Erhalt einer Mandantenerklärung hingegen auf die Breitensuche der Spezialisten (3,23 ggü. 3,00;  $p = 0,24$ ). Somit ist  $H_{II.2.2}$  abzulehnen.<sup>3</sup>

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Mann-Whitney-U-Test.

<sup>3</sup> Darüber hinaus konnte kein signifikanter Einfluss der Mandantenerklärung auf den Umfang der Suche und die Tiefensuche in Bezug auf Kat. 2, 3, 4 und 5 festgestellt werden.

|                                                         | Mit<br>Mandantenerklärung<br>n=13 | Ohne<br>Mandantenerklärung<br>n=12 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| Breite<br>$\mu$<br>[ $\sigma$ ]                         | 3,23<br>[1,01]                    | 3,00<br>[1,35]                     | 0,24                                   |
| Kat. 1<br>(Mandantenerklärung)<br>$\mu$<br>[ $\sigma$ ] | 1,77<br>[0,93]                    | 1,25<br>[1,06]                     | 0,09                                   |

Tabelle 26: Breite der Suche und Tiefe der Kat. 1-Tests der Spezialisten (Fehler-Version)

#### 4.4.4.2.3.2 Nicht-Fehler-Version

##### *Nicht-Spezialisten*

Wie in Tabelle 27 zu erkennen ist, führen Nicht-Spezialisten, denen eine Mandantenerklärung mitgeteilt wurde, durchschnittlich mehr Kat. 1-Tests durch als Nicht-Spezialisten ohne Mandantenerklärung (1,83 ggü. 1,27;  $p = 0,11$ ). Obwohl die Ergebnisse in die vermutete Richtung gehen, sind die Unterschiede jedoch nicht signifikant, weshalb  $H_{II.1.1}$  abzulehnen ist. Weiterhin zeigen die Ergebnisse, dass kein signifikanter Einfluss der Mandantenerklärung auf die Breitensuche besteht (3,08 ggü. 3,27;  $p = 0,21$ ). Demnach ist auch  $H_{II.2.1}$  abzulehnen.

|                                                         | Mit<br>Mandantenerklärung<br>n=12 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| Breite<br>$\mu$<br>[ $\sigma$ ]                         | 3,08<br>[0,79]                    | 3,27<br>[1,01]                     | 0,21                                   |
| Kat. 1<br>(Mandantenerklärung)<br>$\mu$<br>[ $\sigma$ ] | 1,83<br>[1,11]                    | 1,27<br>[1,10]                     | 0,11                                   |

Tabelle 27: Breite der Suche und Tiefe der Kat. 1-Tests der Nicht-Spezialisten (Nicht-Fehler-Version)

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Mann-Whitney-U-Test.

Vergleichbar zur Informationssuche der Nicht-Spezialisten in der Fehler-Version<sup>1</sup> zeigt Tabelle 28, dass Nicht-Spezialisten mit Mandantenerklärung auch hier signifikant mehr Kat. 5-Tests (1,33 ggü. 0,64;  $p = 0,006$ ) und signifikant weniger Kat. 4-Tests (0,00 ggü. 0,18;  $p = 0,06$ ) durchführten.<sup>2, 3</sup>

|                                                                     | Mit<br>Mandantenerklärung<br>n=12 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>4</sup><br>(1-seitig) |
|---------------------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Kat. 4<br/>(Nicht-Fehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,00<br>[0,00]                    | 0,18<br>[0,40]                     | 0,06                                   |
| <b>Kat. 5<br/>(Fehler in den Vorräten)</b><br>$\mu$<br>[ $\sigma$ ] | 1,33<br>[0,65]                    | 0,64<br>[0,51]                     | 0,006                                  |

Tabelle 28: Tiefe der Kat. 4 & 5-Tests der Nicht-Spezialisten (Nicht-Fehler-Version)

### *Spezialisten*

Entgegen der Erwartung von  $H_{II.1.2}$  zeigt Tabelle 29, dass Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhielten, fast identisch viele Tests zur Überprüfung der Mandantenerklärung durchführten wie Spezialisten ohne Mandantenerklärung (1,70 ggü. 1,72;  $p = 1,00$ ).  $H_{II.1.2}$  ist demzufolge abzulehnen. Darüber hinaus geben die Ergebnisse zu erkennen, dass Spezialisten auch keine geringere Breitensuche praktizierten, wenn ihnen eine Mandantenerklärung zur Verfügung gestellt wurde ( $p = 0,11$ ). Somit ist  $H_{II.2.2}$  ebenfalls abzulehnen. Vielmehr weisen die Ergebnisse darauf hin, dass Spezialisten mit Mandantenerklärung ein breiteres Suchverhalten zeigten als Spezialisten ohne Mandantenerklärung (3,60 ggü. 2,91). Diese Beobachtung könnte ähnlich wie die verstärkte anfängliche Generierung von Fehlerhypothesen der Spezialisten auf die erhöhte Risikowahrnehmung zurückzuführen sein, wenn der Mandant die Abweichung mit einer Nicht-Fehlerursache begründet.<sup>5</sup>

<sup>1</sup> Vgl. hierzu Kap. 4.4.4.2.3.1.

<sup>2</sup> Zu einer möglichen Begründung dieser Beobachtung soll auf die Ausführungen unter Kap. 4.4.4.2.3.1. verwiesen werden.

<sup>3</sup> Kein Einfluss konnte hingegen auf den Suchumfang sowie auf die Tiefensuche in Bezug auf Kat. 2 und 3 festgestellt werden.

<sup>4</sup> Mann-Whitney-U-Test.

<sup>5</sup> Vgl. hierzu die Ausführungen zum „inversen“ spreading effect unter Kap. 4.4.3.2.2.

|                                                                       | Mit<br>Mandantenerklärung<br>n=10 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>1</sup><br>(1-seitig) |
|-----------------------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Breite</b><br>$\mu$<br>[ $\sigma$ ]                                | 3,60<br>[0,84]                    | 2,91<br>[1,14]                     | 0,11                                   |
| <b>Kat. 1</b><br><b>(Mandantenerklärung)</b><br>$\mu$<br>[ $\sigma$ ] | 1,70<br>[0,67]                    | 1,72<br>[0,79]                     | 1,00                                   |

Tabelle 29: Breite der Suche und Tiefe der Kat. 1-Tests der Spezialisten (Nicht-Fehler-Version)

Weiterhin fällt auf, dass Spezialisten mit Mandantenerklärung auf einem 10%-Signifikanzniveau wesentlich mehr Kat. 4-Tests (0,20 ggü. 0,00;  $p = 0,06$ ) und auf einem 5%-Signifikanzniveau wesentlich mehr Kat. 5-Tests (1,50 ggü. 0,73;  $p = 0,02$ ) durchgeführt haben (siehe Tabelle 30). Letzteres kann ebenso wie die oben festgestellte verstärkte Breitensuche von Spezialisten mit Mandantenerklärung auf die erhöhte Risikowahrnehmung zurückgeführt werden, wenn der Mandant die Abweichung mit einer Nicht-Fehlerursache begründet.<sup>2</sup> Ersteres könnte möglicherweise auf die im Rahmen der nachträglichen Hypothesengenerierung von Spezialisten festgestellt verstärkte Generierung von Nicht-Fehlerhypothesen zurückgeführt werden,<sup>3</sup> die im Rahmen der Informationssuche eine verstärkte Überprüfung von Nicht-Fehlerhypothesen bedingten. Insgesamt scheint die vermehrte Durchführung von Kat. 4- und Kat. 5-Tests zu einem erhöhtem Suchumfang der Spezialisten geführt zu haben. Spezialisten mit Mandantenerklärung führten auf einem 10%-Signifikanzniveau wesentlich mehr Tests durch (5,80 ggü. 4,82;  $p = 0,09$ ) und benötigten hierzu wesentlich mehr Prüfungszeit (16,90 Std. ggü. 13,36 Std.;  $p = 0,06$ ) als Spezialisten ohne Mandantenerklärung.

---

<sup>1</sup> Mann-Whitney-U-Test.

<sup>2</sup> Vgl. hierzu die Ausführungen zum „inversen“ spreading effect unter Kap. 4.4.3.2.2.

<sup>3</sup> Vgl. hierzu Kap. 4.4.3.3.2.2.

|                                                                     | Mit<br>Mandantenerklärung<br>n=10 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Zeit</b><br>$\mu$<br>[ $\sigma$ ]                                | 16,90<br>[3,89]                   | 13,36<br>[5,65]                    | 0,06                                   |
| <b>Anzahl</b><br>$\mu$<br>[ $\sigma$ ]                              | 5,80<br>[1,23]                    | 4,82<br>[1,89]                     | 0,09                                   |
| <b>Kat. 4<br/>(Nicht-Fehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,20<br>[0,42]                    | 0,00<br>[0,00]                     | 0,06                                   |
| <b>Kat. 5<br/>(Fehler in den Vorräten)</b><br>$\mu$<br>[ $\sigma$ ] | 1,50<br>[0,85]                    | 0,73<br>[0,79]                     | 0,02                                   |

Tabelle 30: Umfang der Suche sowie Tiefe der Kat. 4- und Kat. 5-Tests der Spezialisten (Nicht-Fehler-Version)

#### 4.4.4.2.4 Test von $FF_{II}$

##### 4.4.4.2.4.1 Fehler-Version

Stellt man die Ergebnisse differenziert nach niedrigem Prüfungsrisiko auf der einen Seite und hohem Prüfungsrisiko auf der anderen Seite gegenüber (siehe Tabelle 31), so kann festgestellt werden, dass der Erhalt einer Mandantenerklärung bei geringem Prüfungsrisiko zu durchschnittlich 1,64 durchgeführten Kat. 1-Test führt gegenüber 1,18 bei erhöhtem Prüfungsrisiko ( $p = 0,12$ ). Die Unterschiede können hier nicht als signifikant bezeichnet werden. Berücksichtigt man jedoch, dass bei erhöhtem Risiko beinahe keine Unterschiede zwischen Prüfern mit und Prüfern ohne Mandantenerklärung bestehen (1,64 ggü 1,60;  $p = 1,00$ ), so ist dennoch darauf aufmerksam zu machen, dass die Unterschiede zwischen dem Einfluss einer Mandantenerklärung bei geringem Risiko einerseits und erhöhtem Risiko andererseits nicht unbeachtet bleiben dürfen. Es besteht zumindest der Verdacht, dass Prüfer mit Mandantenerklärung bei geringem Prüfungsrisiko dazu neigen, vermehrt Tests durchzuführen, die auf die Beurteilung dieser vererbten Hypothese gerichtet sind, und insofern konfirmatorische Prozesse offenbaren.

---

<sup>1</sup> Mann-Whitney-U-Test.



Betrachtet man hingegen den Einfluss einer Mandantenerklärung auf die Breite der Suchstrategie, so fällt auf, dass diesbezüglich lediglich geringfügige Unterschiede in Abhängigkeit des Prüfungsrisikos zu verzeichnen sind (3,45 ggü. 3,18;  $p = 0,73$  bei geringem Prüfungsrisiko bzw. 3,18 ggü. 3,00;  $p = 0,68$  bei erhöhtem Prüfungsrisiko).<sup>1, 2</sup>

Hinsichtlich der Tiefe des Suchverhaltens fällt auf, dass die Prüfer bei geringem Prüfungsrisiko auf einem 10%-Signifikanzniveau wesentlich häufiger Tests durchführten, die auf die Überprüfung des tatsächlichen Abweichungsgrunds (Kat. 2) abzielten, wenn ihnen eine Mandantenerklärung bereitgestellt wurde (2,18 ggü. 1,55;  $p = 0,07$ <sup>3</sup>). Ein ähnlicher Einfluss der Mandantenerklärung konnte bei erhöhtem Prüfungsrisiko ebenfalls nachgewiesen werden (siehe Tabelle 31). Prüfer ohne Mandantenerklärung führten auf einem 10%-Signifikanzniveau wesentlich mehr Kat. 2-Tests durch (2,00 ggü. 1,09;  $p = 0,06$ <sup>4</sup>).<sup>5</sup> Allerdings wäre aufgrund theoretischer Überlegungen<sup>6</sup> zu vermuten, dass der Erhalt einer Nicht-Fehlererklärung des Mandanten zu einer verringerten Durchführung von Tests führt, die auf eine Fehlerursache abzielen. Hier weisen die Ergebnisse jedoch in die entgegengesetzte Richtung.<sup>1</sup>

---

<sup>1</sup> Jeweils Mann-Whitney-U-Test unter Berücksichtigung der 2-seitigen Signifikanz.

<sup>2</sup> Keine wesentlichen Auswirkungen hat das Prüfungsrisiko weiterhin auf den Umfang der Suchstrategie (5,82 ggü. 5,00;  $p = 0,51$  bzw. 5,00 ggü. 5,40;  $p = 0,54$ ).

<sup>3</sup> Mann-Whitney-U-Test unter Berücksichtigung der 1-seitigen Signifikanz.

<sup>4</sup> Mann-Whitney-U-Test (1-seitige Signifikanz).

<sup>5</sup> Dieses entspricht durchaus den Überlegungen, die in  $H_{II.2.1}$  und  $H_{II.2.2}$  zum Ausdruck kommen, wonach davon ausgegangen wird, dass Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, weniger Prüfungsnachweise einholen, die auf die Beurteilung anderer potenzieller Abweichungsursachen gerichtet sind.

<sup>6</sup> Vgl. hierzu insbesondere die Herleitung von  $H_{II.2.1}$  und  $H_{II.2.2}$ .

|                                                                           | Mit<br>Mandantenerklärung |                 | Ohne<br>Mandantenerklärung |                 |
|---------------------------------------------------------------------------|---------------------------|-----------------|----------------------------|-----------------|
|                                                                           | Niedrig<br>n=11           | Hoch<br>n=11    | Niedrig<br>n=11            | Hoch<br>n=10    |
| <b>Zeit</b><br>$\mu$<br>[ $\sigma$ ]                                      | 15,95<br>[3,49]           | 14,55<br>[4,45] | 13,95<br>[5,80]            | 15,35<br>[4,21] |
| <b>Anzahl</b><br>$\mu$<br>[ $\sigma$ ]                                    | 5,82<br>[1,33]            | 5,00<br>[1,55]  | 5,00<br>[1,84]             | 5,40<br>[1,58]  |
| <b>Breite</b><br>$\mu$<br>[ $\sigma$ ]                                    | 3,45<br>[0,69]            | 3,18<br>[0,98]  | 3,18<br>[1,40]             | 3,00<br>[1,05]  |
| <b>Kat. 1</b><br><b>(Mandantenerklärung)</b><br>$\mu$<br>[ $\sigma$ ]     | 1,64<br>[0,92]            | 1,64<br>[1,03]  | 1,18<br>[1,08]             | 1,60<br>[1,07]  |
| <b>Kat. 2</b><br><b>(Korrektur Grund)</b><br>$\mu$<br>[ $\sigma$ ]        | 2,18<br>[1,25]            | 1,09<br>[0,83]  | 1,55<br>[0,69]             | 2,00<br>[1,33]  |
| <b>Kat. 3</b><br><b>(Umsatzfehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,91<br>[0,83]            | 1,45<br>[0,82]  | 0,83<br>[0,70]             | 0,90<br>[0,74]  |
| <b>Kat. 4</b><br><b>(Nicht-Fehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,09<br>[0,30]            | 0,00<br>[0,00]  | 0,27<br>[0,47]             | 0,10<br>[0,32]  |
| <b>Kat. 5</b><br><b>(Fehler in den Vorräten)</b><br>$\mu$<br>[ $\sigma$ ] | 1,00<br>[0,45]            | 0,82<br>[0,60]  | 1,09<br>[1,14]             | 0,80<br>[1,23]  |

Tabelle 31: Angewandte Suchstrategie der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

<sup>1</sup> Da vermutet wird, dass diese Beobachtung auf das Suchverhalten der Nicht-Spezialisten zurückzuführen ist, sei an diese Stelle auf die Ausführungen unter Kap. 4.4.4.2.3.1 verwiesen.

Bei erhöhtem Prüfungsrisiko zeigt Tabelle 31, dass Probanden, die eine Nicht-Fehlererklärung des Mandanten erhalten, auf einem 5%-Signifikanzniveau wesentlich häufiger Tests hinsichtlich Kat. 3 (Umsatzfehler) durchführen (0,90 ggü. 1,45;  $p = 0,05^1$ ). Ein derartiger Einfluss der Mandantenerklärung konnte bei geringem Prüfungsrisiko hingegen nicht festgestellt werden (0,91 ggü. 0,83). Hier ist zu vermuten, dass der Erhalt einer Mandantenerklärung unter Berücksichtigung eines erhöhten Prüfungsrisikos – hier sei noch mal darauf hingewiesen, dass der Mandant die festgestellte Abweichung auf die gestiegene Nachfrage und somit auf eine Nicht-Fehlerursache im Bereich der Umsatzerlöse zurückführt – die Aufmerksamkeit des Prüfers auf die Umsatzerlöse richtete. Geht man davon aus, dass Prüfer ihr Wissen anhand von Transaktionskreisen strukturieren<sup>2</sup>, so könnte diese Beobachtung ebenfalls auf einen spreading effect zurückgeführt werden, der sich allerdings nicht auf Nicht-Fehlerhypothesen beschränkt, sondern sich sowohl auf Fehler- als auch auf Nicht-Fehlererklärungen innerhalb eines bestimmten Transaktionskreises bezieht. Insofern könnte diese Beobachtung auch als *transaktionskreisgebundener spreading effect* bezeichnet werden.

#### 4.4.4.2.4.2 Nicht-Fehler-Version

Wie Tabelle 32 zeigt, führten Prüfer mit Mandantenerklärung auf einem 5%-Signifikanzniveau wesentlich mehr Kat. 1-Tests durch (1,75 ggü. 1,25;  $p = 0,05^3$ ), wenn ein erhöhtes Prüfungsrisiko zu verzeichnen war. Bei geringem Prüfungsrisiko war dieser Unterschied hingegen nicht evident (1,80 ggü. 1,80;  $p = 1,00^4$ ). Demzufolge ist zu konstatieren, dass vermutlich ein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss einer Mandantenerklärung auf die Einholung von Informationen zur Beurteilung der Mandantenerklärung besteht. Erhöhtes Prüfungsrisiko scheint somit ein Verstärker einer auf die Mandantenerklärung bezogenen Tiefensuche zu sein. Kein Zusammenhang konnte hingegen zwischen dem Einfluss einer Mandantenerklärung auf die Breitensuche und dem Prüfungsrisiko nachgewiesen werden. Hier konnten weder bei erhöhtem noch bei geringem Prüfungsrisiko signifikante Unterschiede zwischen Prüfern mit und Prüfern ohne Mandantenerklärung nachgewiesen werden (3,17 ggü.

---

<sup>1</sup> Mann-Whitney-U-Test (1-seitige Signifikanz).

<sup>2</sup> Vgl. hierzu Kap. 3.5.3.3.1.2.1 i.V.m. 3.5.3.3.2.1.2.2 sowie 3.2.2.3.4.2.2.

<sup>3</sup> Mann-Whitney-U-Test (1-seitige Signifikanz). Dieses Verhalten stimmt durchaus mit der in  $H_{II,1,1}$  und  $H_{II,1,2}$  zum Ausdruck kommenden Vermutung überein.

<sup>4</sup> Mann-Whitney-U-Test (1-seitige Signifikanz).

3,08;  $p = 0,41$  bei erhöhtem Risiko und 3,50 ggü. 3,10;  $p = 0,22$  bei niedrigem Prüfungsrisiko)<sup>1</sup>.

Bei erhöhtem Prüfungsrisiko ist ferner zu erkennen, dass Prüfer mit Mandantenerklärung im Schnitt 6,17 Tests durchführten und dafür 17,71 Stunden benötigten im Vergleich zu durchschnittlich 5,08 durchgeführten Tests in 14,17 Stunden bei Probanden ohne Mandantenerklärung. Hinsichtlich der Prüfungszeit ist auf einem 5%-Signifikanzniveau ein wesentlicher Einfluss zu verzeichnen ( $p = 0,05^2$ ). Im Gegensatz hierzu zeigt Tabelle 32, dass die Mandantenerklärung bei geringem Prüfungsrisiko weder einen signifikanten Einfluss auf den Umfang der Suche noch auf die Anzahl durchgeführten Test ausübte (15,55 Std. ggü. 13,70 Std;  $p = 0,33$  bzw. 5,40 ggü. 4,90;  $p = 0,50^3$ ). Somit weisen die Ergebnisse darauf hin, dass der Einfluss der Mandantenerklärung auf den Suchumfang durch erhöhtes Prüfungsrisiko verstärkt wird.

Ferner konnte sowohl bei erhöhtem als auch bei niedrigem Prüfungsrisiko ein signifikanter Einfluss der Mandantenerklärung auf die Durchführung von Kat. 5-Tests festgestellt werden. Bei erhöhtem Risiko führten Prüfer mit Mandantenerklärung auf einem 10%-Signifikanzniveau wesentlich mehr Kat. 5-Tests durch (1,50 ggü. 0,83;  $p = 0,06^4$ ). Bei geringem Prüfungsrisiko konnte der Einfluss indes auf einem 1%-Signifikanzniveau nachgewiesen werden (1,30 ggü. 0,50;  $p = 0,005^5$ ).<sup>6</sup>

---

<sup>1</sup> Mann-Whitney-U-Test (1-seitige Signifikanz).

<sup>2</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>3</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>4</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>5</sup> Mann-Whitney-U-Test (2-seitige Signifikanz).

<sup>6</sup> Für eine Erklärung dieser Beobachtung soll auf die Ausführungen unter Kap. 4.4.4.2.3.2 verwiesen werden, da bei Spezialisten einerseits und Nicht-Spezialisten andererseits scheinbar unterschiedliche Ursachen zu diesem Effekt geführt haben.

|                                                                           | Mit<br>Mandantenerklärung |                 | Ohne<br>Mandantenerklärung |                 |
|---------------------------------------------------------------------------|---------------------------|-----------------|----------------------------|-----------------|
|                                                                           | Niedrig<br>n=10           | Hoch<br>n=12    | Niedrig<br>n=12            | Hoch<br>n=10    |
| <b>Zeit</b><br>$\mu$<br>[ $\sigma$ ]                                      | 15,55<br>[5,33]           | 17,71<br>[3,06] | 13,70<br>[5,62]            | 14,17<br>[5,20] |
| <b>Anzahl</b><br>$\mu$<br>[ $\sigma$ ]                                    | 5,40<br>[1,71]            | 6,17<br>[1,03]  | 4,90<br>[1,974]            | 5,08<br>[1,73]  |
| <b>Breite</b><br>$\mu$<br>[ $\sigma$ ]                                    | 3,50<br>[0,97]            | 3,17<br>[0,72]  | 3,10<br>[1,10]             | 3,08<br>[1,08]  |
| <b>Kat. 1</b><br><b>(Mandantenerklärung)</b><br>$\mu$<br>[ $\sigma$ ]     | 1,80<br>[1,14]            | 1,75<br>[0,75]  | 1,80<br>[1,03]             | 1,25<br>[0,87]  |
| <b>Kat. 2</b><br><b>(Korrektur Grund)</b><br>$\mu$<br>[ $\sigma$ ]        | 1,10<br>[0,74]            | 2,00<br>[1,28]  | 1,60<br>[1,35]             | 1,83<br>[1,40]  |
| <b>Kat. 3</b><br><b>(Umsatzfehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 1,10<br>[0,88]            | 0,83<br>[1,03]  | 0,90<br>[0,57]             | 1,08<br>[1,00]  |
| <b>Kat. 4</b><br><b>(Nicht-Fehler)</b><br>$\mu$<br>[ $\sigma$ ]           | 0,10<br>[0,32]            | 0,08<br>[0,29]  | 0,10<br>[0,32]             | 0,08<br>[0,29]  |
| <b>Kat. 5</b><br><b>(Fehler in den Vorräten)</b><br>$\mu$<br>[ $\sigma$ ] | 1,30<br>[0,48]            | 1,50<br>[0,90]  | 0,50<br>[0,53]             | 0,83<br>[0,72]  |

Tabelle 32: Angewandte Suchstrategie der Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version)

#### 4.4.4.3 Der Einfluss auf die Interpretation von Prüfungsnachweisen

##### 4.4.4.3.1 Vorbemerkungen

Zur Beurteilung des Einflusses einer Nicht-Fehlererklärung des Mandanten auf die Interpretation von Prüfungsnachweisen hinsichtlich der Mandantenerklärung einerseits und hinsichtlich des tatsächlichen Abweichungsgrunds andererseits wurde untersucht, ob und in welche Richtung der jeweils zugeordnete Rang nach Kenntnisnahme eines auf die Beurteilung der Mandantenerklärung bzw. des tatsächlichen Abweichungsgrunds gerichteten Testergebnisses modifiziert wurde.<sup>1, 2</sup> Hierzu wurden zunächst alle von den Probanden ausgewählten Kat. 1-Tests (Mandantenerklärung) und Kat. 2-Tests (korrekter Grund) selektiert. Anschließend wurde überprüft, ob die Testergebnisse als bestätigend oder nicht bestätigend interpretiert wurden. Dabei wurde die Interpretation eines Testergebnisses als bestätigend gewertet, wenn ein Rang  $< 1$  nach oben angepasst oder ein Rang  $= 1$  beibehalten wurde. Eine Interpretation wurde hingegen als nicht bestätigend klassifiziert, wenn ein Rang  $\leq 1$  nach unten korrigiert oder ein Rang  $< 1$  beibehalten wurde.

##### 4.4.4.3.2 Test von $H_{III.1}$ und $H_{III.2}$

###### 4.4.4.3.2.1 Fehler-Version

Tabelle 33 zeigt, dass Prüfer mit Mandantenerklärung Tests, die auf die Beurteilung dieser Erklärung gerichtet waren, in 47,2% der Fälle als bestätigend interpretierten im Vergleich zu 37,9% bei Prüfern ohne Mandantenerklärung. Obwohl diese Ergebnisse in die erwartete Richtung weisen, sind die Unterschiede nicht signifikant ( $p = 0,23$ ) und  $H_{III.1}$  mithin abzulehnen. Ein durch den Erhalt einer Mandantenerklärung ausgelöster confirmation bias konnte mithin nicht nachgewiesen werden.

---

<sup>1</sup> Die Probanden wurden aufgefordert, ihre generierten Hypothesen nach abnehmender Eintrittswahrscheinlichkeit in eine Rangordnung von 1 bis N zu bringen. Vgl. hierzu auch die Ausführung unter Kap. 4.3.2.

<sup>2</sup> Da bei der Nicht-Fehler-Version die Mandantenerklärung zugleich die tatsächliche Abweichungsursache ist, kann hier lediglich auf die Interpretation der Testergebnisse zur Mandantenerklärung eingegangen werden.

|                                 | Mit<br>Mandantenerklärung<br>n=36 | Ohne<br>Mandantenerklärung<br>n=29 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 47,2                              | 37,9                               | 0,23                                   |
| <b>Nicht bestätigend</b><br>[%] | 52,8                              | 62,1                               |                                        |

Tabelle 33: : Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer (Fehler-Version)

Wie Tabelle 34 zu entnehmen ist, interpretierten Prüfer mit Mandantenerklärung Tests, die auf die Beurteilung der tatsächlichen Abweichungsursache gerichtet waren, zu 41,7% als bestätigend im Vergleich zu 35,1% bei Prüfern ohne Mandantenerklärung. Da der Unterschied nicht signifikant ist ( $p = 0,28$ ), ist  $H_{III,2}$  abzulehnen.

|                                 | Mit<br>Mandantenerklärung<br>n=36 | Ohne<br>Mandantenerklärung<br>n=37 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 41,7                              | 35,1                               | 0,28                                   |
| <b>Nicht bestätigend</b><br>[%] | 58,3                              | 64,9                               |                                        |

Tabelle 34: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Prüfer (Fehler-Version)

#### 4.4.4.3.2.2 Nicht-Fehler-Version

Prüfer mit Mandantenerklärung interpretierten Tests zur Beurteilung dieser Erklärung zu 69,2% als bestätigend im Vergleich zu 51,5% bei Prüfern ohne Mandantenerklärung (siehe Tabelle 35). Auf einem 10%-Signifikanzniveau sind diese Unterschiede zwischen den beiden Gruppen signifikant und  $H_{III,1}$  zu bestätigen.

---

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

|                                 | Mit<br>Mandantenerklärung<br>n=39 | Ohne<br>Mandantenerklärung<br>n=33 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 69,2                              | 51,5                               | 0,06                                   |
| <b>Nicht bestätigend</b><br>[%] | 30,8                              | 48,5                               |                                        |

Tabelle 35: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer (Nicht-Fehler-Version)

#### 4.4.4.3.3 Test von $H_{III.1.1}$ bis $H_{III.2.2}$

##### 4.4.4.3.3.1 Fehler-Version

###### *Nicht-Spezialisten*

Wie in Tabelle 36 zu erkennen ist, interpretierten Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhielten, die Prüfungsergebnisse, die sich auf die Mandantenerklärung beziehen, häufiger als bestätigend als Prüfer ohne Mandantenerklärung (69,2% ggü. 50%;  $p = 0,15$ ). Obwohl die Ergebnisse somit in die von  $H_{III.1.1}$  erwartete Richtung weisen, ist der Unterschied jedoch nicht signifikant.  $H_{III.1.1}$  ist folglich abzulehnen. Durch den Erhalt einer Mandantenerklärung ausgelöste konfirmatorische Prozesse waren im Rahmen dieser Studie folglich nicht evident.

|                                 | Mit<br>Mandantenerklärung<br>n=13 | Ohne<br>Mandantenerklärung<br>n=14 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 69,2                              | 50,0                               | 0,15                                   |
| <b>Nicht bestätigend</b><br>[%] | 30,8                              | 50,0                               |                                        |

Tabelle 36: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Nicht-Spezialisten (Fehler-Version)

Allerdings ist Tabelle 37 zu entnehmen, dass der Erhalt einer Mandantenerklärung einen signifikanten Einfluss auf die Interpretation von Ergebnissen hat, die sich auf die Beurteilung der tatsächlichen Abweichungsursache beziehen. Nicht-Spezialisten mit Man-

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.



mandantenerklärung interpretierten diese auf einem 5%-Signifikanzniveau wesentlich häufiger als bestätigend und somit richtig. Allerdings gehen diese Ergebnisse in eine andere als durch  $H_{III.2.1}$  vermutete Richtung.  $H_{III.2.1}$  ist somit abzulehnen. Demnach weisen die Befunde darauf hin, dass der Erhalt einer Mandantenerklärung auch bei der Interpretation von Ergebnissen hinsichtlich des tatsächlichen Abweichungsgrunds nicht dazu führt, dass Hinweise, die die Gültigkeit der Mandantenerklärung bezweifeln lassen, falsch interpretiert bzw. „heruntergespielt“ werden.

Fraglich ist allerdings, weshalb die Ergebnisse in die entgegengesetzte Richtung weisen. Hierzu kann erneut auf die bei Nicht-Spezialisten durch Erhalt einer Nicht-Fehlererklärung erhöhte Risikowahrnehmung zurückgegriffen werden.<sup>1</sup> Dieses wahrgenommene erhöhte Prüfungsrisiko könnte zu einer gesteigerten Aufmerksamkeit der Nicht-Spezialisten bezüglich Hinweisen geführt haben, die gegen die Gültigkeit der „vererbten“ Nicht-Fehlererklärung sprechen.

|                                 | Mit<br>Mandantenerklärung<br>n=14 | Ohne<br>Mandantenerklärung<br>n=16 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 57,1                              | 25,0                               | 0,04                                   |
| <b>Nicht bestätigend</b><br>[%] | 42,9                              | 75,0                               |                                        |

Tabelle 37: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Nicht-Spezialisten (Fehler-Version)

### *Spezialisten*

Wie Tabelle 38 zeigt, interpretierten Spezialisten, die eine Nicht-Fehlererklärung des Mandanten zur Verfügung gestellt bekamen, die Ergebnisse zur Beurteilung dieser Erklärung nicht signifikant häufiger als bestätigend (34,8% ggü. 26,7%;  $p = 0,30$ ). Folglich ist  $H_{III.1.2}$  abzulehnen. Im Rahmen dieser Studie lassen Spezialisten bei Erhalt einer Mandantenerklärung keine confirmatorische Prozesse erkennen.

<sup>1</sup> Vgl. hierzu ausführlich die Ausführungen unter Kap. 4.4.4.2.3.1.

<sup>2</sup> Chi-Quadrat-Test.

|                                 | Mit<br>Mandantenerklärung<br>n=23 | Ohne<br>Mandantenerklärung<br>n=15 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 34,8                              | 26,7                               | 0,30                                   |
| <b>Nicht bestätigend</b><br>[%] | 65,2                              | 73,3                               |                                        |

Tabelle 38: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Spezialisten (Fehler-Version)

In Bezug auf die Interpretation von Prüfungsergebnissen, die auf die tatsächliche Abweichungsursache gerichtet sind, weist Tabelle 39 ebenfalls darauf hin, dass keine signifikanten Unterschiede in Abhängigkeit des Erhalts einer Nicht-Fehlererklärung des Mandanten vorliegen. Spezialisten mit Mandantenerklärung interpretieren Tests hinsichtlich des tatsächlichen Grunds zu 31,8% als bestätigend gegenüber 42,9% bei Spezialisten ohne Mandantenerklärung ( $p = 0,23$ ).  $H_{III.2.2}$  ist somit ebenfalls abzulehnen. Spezialisten, denen eine Nicht-Fehlererklärung vererbt wurde, konnten folglich keine konfirmatorischen Prozesse nachgewiesen werden.

|                                 | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=21 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend</b><br>[%]       | 31,8                              | 42,9                               | 0,23                                   |
| <b>Nicht bestätigend</b><br>[%] | 68,2                              | 57,1                               |                                        |

Tabelle 39: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Spezialisten (Fehler-Version)

#### 4.4.4.3.2 Nicht-Fehler-Version<sup>1</sup>

##### *Nicht-Spezialisten*

Ebenso wie im Rahmen der Fehler-Version weisen die Ergebnisse in Tabelle 40 darauf hin, dass Nicht-Spezialisten auch im Rahmen der Nicht-Fehler-Version keine konfirmatorischen Prozesse nachzuweisen sind, wenn ihnen eine Nicht-Fehlererklärung des

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

Mandanten mitgeteilt wird. Nicht-Spezialisten interpretierten Prüfungsergebnisse, die auf die Überprüfung der Korrektheit der Mandantenerklärung gerichtet sind, zu 68,2% als bestätigend im Vergleich zu 57,1% bei Nicht-Spezialisten ohne Mandantenerklärung ( $p = 0,25$ ). Somit ist  $H_{III.1.1}$  auch hier abzulehnen.

|                                  | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=14 | Signifikanz <sup>2</sup><br>(1-seitig) |
|----------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend<br/>[%]</b>       | 68,2                              | 57,1                               | 0,25                                   |
| <b>Nicht bestätigend<br/>[%]</b> | 31,8                              | 42,9                               |                                        |

Tabelle 40: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Nicht-Spezialisten (Nicht-Fehler-Version)

### *Spezialisten*

Tabelle 41 zeigt, dass die Vererbung einer Nicht-Fehlererklärung auf einem 10%-Signifikanzniveau einen wesentlichen Einfluss auf die Interpretation von Ergebnissen hat, die auf die Beurteilung der Mandantenerklärung gerichtet sind. Spezialisten mit Mandantenerklärung interpretierten auf die Mandantenerklärung gerichtete Prüfungsergebnisse signifikant häufiger als bestätigend (70,6% ggü. 47,4%;  $p = 0,08$ ). Somit ist  $H_{III.1.2}$  zu bestätigen. Im Gegensatz zur Fehler-Version konnten Spezialisten bei Erhalt einer Mandantenerklärung confirmatorische Prozesse nachgewiesen werden.

|                                  | Mit<br>Mandantenerklärung<br>n=17 | Ohne<br>Mandantenerklärung<br>n=19 | Signifikanz <sup>1</sup><br>(1-seitig) |
|----------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Bestätigend<br/>[%]</b>       | 70,6                              | 47,4                               | 0,08                                   |
| <b>Nicht bestätigend<br/>[%]</b> | 29,4                              | 52,6                               |                                        |

Tabelle 41: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Spezialisten (Nicht-Fehler-Version)

<sup>1</sup> Da es sich im Rahmen der Nicht-Fehler-Version bei der Mandantenerklärung um die tatsächliche Abweichungsursache handelt, können hier lediglich  $H_{III.1.1}$  und  $H_{III.1.2}$  überprüft werden.

<sup>2</sup> Chi-Quadrat-Test.

#### 4.4.4.3.4 Test von $FF_{III}$

##### 4.4.4.3.4.1 Fehler-Version

Unter Berücksichtigung des zugrunde liegenden Prüfungsrisikos weist Tabelle 42 darauf hin, dass Prüfer mit Mandantenerklärung die Testergebnisse hinsichtlich des tatsächlichen Abweichungsgrunds bei geringem Prüfungsrisiko zu 37,5% als bestätigend interpretierten im Vergleich zu 47,1% bei Probanden ohne Mandantenerklärung. Es fällt auf, dass diese Ergebnisse in die durch  $H_{III.2.1}$  und  $H_{III.2.2}$  vermutete Richtung weisen. Dieser Unterschied ist jedoch nicht signifikant ( $p = 0,27$ ).<sup>2</sup> Anders verhält es sich hingegen bei erhöhtem Prüfungsrisiko. Hier interpretierten Prüfer mit Mandantenerklärung die Testergebnisse zu 50% als bestätigend. Demgegenüber interpretierten Prüfer ohne Mandantenerklärung die Testergebnisse zu 25% als bestätigend. Auf einem 10%-Signifikanzniveau ist dieser Unterschied signifikant.<sup>3</sup> Jedoch weisen diese Ergebnisse eine entgegengesetzte Tendenz zu der durch  $H_{III.2.1}$  und  $H_{III.2.2}$  vermuteten Richtung auf.<sup>4</sup> Die Ergebnisse lassen also vermuten, dass das durch den Erhalt einer Mandantenerklärung erwartete Auftreten konfirmatorischer Prozesse durch erhöhtes Risiko nicht nur vollständig eliminiert wird. Vielmehr erscheint es möglich, dass der Erhalt einer Mandantenerklärung unter erhöhtem Prüfungsrisiko die Aufmerksamkeit derart „schärft“, dass Hinweise, die auf das Vorliegen einer Fehlerursache bzw. die Nicht-Korrektheit der Mandantenerklärung signifikant häufiger erkannt werden.

|                              | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|------------------------------|------------------------|--------------|-------------------------|--------------|
|                              | Niedrig<br>n=24        | Hoch<br>n=12 | Niedrig<br>n=17         | Hoch<br>n=20 |
| <b>Bestätigend [%]</b>       | 37,5                   | 50,0         | 47,1                    | 25,0         |
| <b>Nicht bestätigend [%]</b> | 62,5                   | 50,0         | 52,9                    | 75,0         |

Tabelle 42: Interpretation der Prüfungsergebnisse hinsichtlich der tatsächlichen Abweichungsursache durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>3</sup>  $p = 0,09$  (Chi-Quadrat-Test; 1-seitige Signifikanz).

<sup>4</sup> Es ist zu vermuten, dass dieser Effekt ausschließlich auf das Verhalten der Nicht-Spezialisten zurückzuführen ist. Siehe hierzu Kap. 4.4.4.2.1.1.

Kein signifikanter Zusammenhang konnte hingegen zwischen dem Prüfungsrisiko und dem Einfluss einer Mandantenerklärung auf die Interpretation von Informationen, die auf die Beurteilung der Mandantenerklärung gerichtet waren, festgestellt werden (siehe Tabelle 43). Unter niedrigem Risiko interpretierten Prüfer mit Mandantenerklärung die Ergebnisse zu 61,1% als bestätigend im Vergleich zu 53,8% bei Prüfern ohne Mandantenerklärung ( $p = 0,34$ ). Ähnlich verhielt es sich unter Berücksichtigung eines erhöhten Prüfungsrisikos (33,3% ggü. 25,0%).<sup>1</sup>

|                              | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|------------------------------|------------------------|--------------|-------------------------|--------------|
|                              | Niedrig<br>n=18        | Hoch<br>n=18 | Niedrig<br>n=13         | Hoch<br>n=16 |
| <b>Bestätigend [%]</b>       | 61,1                   | 33,3         | 53,8                    | 25,0         |
| <b>Nicht bestätigend [%]</b> | 38,9                   | 66,7         | 46,2                    | 75,0         |

Tabelle 43: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

#### 4.4.4.3.4.2 Nicht-Fehler-Version

Wie Tabelle 44 zu entnehmen ist, interpretierten Prüfer mit Mandantenerklärung die Testergebnisse bei geringem Prüfungsrisiko zu 66,7% als bestätigend im Vergleich zu 44,4% bei Prüfern ohne Mandantenerklärung. Dieser Unterschied kann auf einem 10%-Signifikanzniveau als wesentlich bezeichnet werden.<sup>2</sup> Kein signifikanter Unterschied konnte hingegen bei erhöhtem Prüfungsrisiko festgestellt werden (71,4% ggü. 60,0%;  $p = 0,24$ ).<sup>3</sup> Diese Ergebnisse legen die Vermutung nahe, dass ein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss des Erhalts einer Mandantenerklärung auf die Interpretation von Informationen zur Beurteilung der Mandantenerklärung besteht. Konfirmatorische Prozesse scheinen insbesondere durch geringes Prüfungsrisiko begünstigt zu werden. Erhöhtem Prüfungsrisiko ist im Rahmen dieser Studie indes möglicherweise eine Debiasing-Funktion zuzusprechen, da konfirmatorische Prozesse hier nicht nachgewiesen werden konnten.

<sup>1</sup> Jeweils Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>2</sup>  $p = 0,09$  (Chi-Quadrat-Test; 1-seitige Signifikanz).

<sup>3</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

|                              | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|------------------------------|------------------------|--------------|-------------------------|--------------|
|                              | Niedrig<br>n=18        | Hoch<br>n=21 | Niedrig<br>n=18         | Hoch<br>n=15 |
| <b>Bestätigend [%]</b>       | 66,7                   | 71,4         | 44,4                    | 60,0         |
| <b>Nicht bestätigend [%]</b> | 33,3                   | 28,6         | 55,6                    | 40,0         |

Tabelle 44: Interpretation der Prüfungsergebnisse hinsichtlich der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version)

#### 4.4.5 Hypothesenbewertung<sup>1</sup>

##### 4.4.5.1 Test von $H_{IV.1.1}$

##### 4.4.5.1.1 Fehler-Version

Wie Tabelle 45 zeigt, selektierten Prüfer, denen eine Mandantenerklärung mitgeteilt wurde, diese ähnlich häufig wie Prüfer ohne Mandantenerklärung (36,4% ggü. 33,3%;  $p = 0,42$ ).  $H_{IV.1}$  ist folglich abzulehnen. Ein Einfluss des Erhalts einer Mandantenerklärung auf ihre Auswahlhäufigkeit konnte in der vorliegenden Studie nicht nachgewiesen werden.

|                                                 | Mit Mandantenerklärung<br>n=22 | Ohne Mandantenerklärung<br>n=21 | Signifikanz <sup>2</sup><br>(1-seitig) |
|-------------------------------------------------|--------------------------------|---------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit Mandantenerklärung [%]</b> | 36,4                           | 33,3                            | 0,42                                   |

Tabelle 45: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer (Fehler-Version)

##### 4.4.5.1.2 Nicht-Fehler-Version

Aus Tabelle 46 geht hervor, dass 72,7% der Prüfer mit Mandantenerklärung diese auch als wahrscheinlichste Abweichungsursache selektierten. Demgegenüber wählten Prüfer ohne Mandantenerklärung die entsprechende Ursache nur in 45,5% der Fälle aus ( $p =$

<sup>1</sup> Da die Mandantenerklärung im Rahmen der Nicht-Fehler-Version zugleich die tatsächliche Abweichungsursache ist, stellt die Auswahlhäufigkeit der Mandantenerklärung hier zugleich die Erfolgsquote der Probanden dar.

<sup>2</sup> Chi-Quadrat-Test.

0,03). Da der Unterschied zwischen den Prüfern mit und den Prüfern ohne Mandantenerklärung auf einem 5%-Signifikanzniveau wesentlich ist, ist  $H_{IV.1}$  anzunehmen.

|                                                         | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=22 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 72,7                              | 45,5                               | 0,03                                   |

Tabelle 46: Auswahlhäufigkeit der Mandantenerklärung ohne Berücksichtigung der branchenspezifischen Erfahrung (Nicht-Fehler-Version)

#### 4.4.5.2 Test von $H_{IV.1.1}$ und $H_{IV.1.2}$

##### 4.4.5.2.1 Fehler-Version

###### *Nicht-Spezialisten*

Wie Tabelle 47 entnommen werden kann, weisen Nicht-Spezialisten unabhängig vom Erhalt einer Mandantenerklärung eine identische Auswahlhäufigkeit auf. Somit ist  $H_{IV.1.1}$  abzulehnen. Nicht-Spezialisten konnte bei der Bewertung einer Mandantenerklärung keine signifikante Beeinflussung durch die Vererbung der entsprechenden Hypothese nachgewiesen werden.

|                                                         | Mit<br>Mandantenerklärung<br>n=9 | Ohne<br>Mandantenerklärung<br>n=9 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------------------------------|----------------------------------|-----------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 44,4                             | 44,4                              | 1,00                                   |

Tabelle 47: Auswahlhäufigkeit der Mandantenerklärung durch Nicht-Spezialisten (Fehler-Version)

###### *Spezialisten*

Wie die Ergebnisse in Tabelle 48 zeigen, konnte auch bei Spezialisten kein signifikanter Einfluss der Vererbung einer Mandantenerklärung auf deren Auswahlhäufigkeit nachgewiesen werden. Obwohl die Ergebnisse in die vermutete Richtung weisen, ist der Unterschied nicht signifikant (30,8% ggü. 25,5%;  $p = 0,37$ ). Somit ist auch  $H_{IV.1.2}$  zurückzuweisen.

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

|                                                         | Mit<br>Mandantenerklärung<br>n=13 | Ohne<br>Mandantenerklärung<br>n=12 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 30,8                              | 25,0                               | 0,37                                   |

Tabelle 48: Auswahlhäufigkeit der Mandantenerklärung durch Spezialisten (Fehler-Version)

#### 4.4.5.2.2 Nicht-Fehler-Version

##### *Nicht-Spezialisten*

Entsprechend der in  $H_{IV.1.1}$  enthaltenen Vermutung zeigt Tabelle 49, dass Nicht-Spezialisten mit Mandantenerklärung diese auf einem 5%-Signifikanzniveau wesentlich häufiger als wahrscheinlichste Abweichungsursache auswählten als Nicht-Spezialisten ohne Mandantenerklärung (66,7% ggü. 27,3%;  $p = 0,03$ ). Somit ist  $H_{IV.1.1}$  nicht abzulehnen. Der Erhalt einer Mandantenerklärung übte bei Nicht-Spezialisten einen signifikanten Einfluss auf deren Auswahlhäufigkeit aus.<sup>2</sup>

|                                                         | Mit<br>Mandantenerklärung<br>n=12 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>1</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 66,7                              | 27,3                               | 0,03                                   |

Tabelle 49: Auswahlhäufigkeit der Mandantenerklärung durch Nicht-Spezialisten (Nicht-Fehler-Version)

##### *Spezialisten*

Wie Tabelle 50 zeigt, selektierten Spezialisten die Mandantenerklärung häufiger als wahrscheinlichste Abweichungsursache, wenn ihnen diese vererbt wurde (80,0% ggü. 63,6%;  $p = 0,20$ ). Obwohl die Resultate somit in die von  $H_{IV.1.2}$  vermutete Richtung gehen, sind die Unterschiede nicht signifikant. Folglich ist  $H_{IV.1.2}$  nicht anzunehmen. Die Auswahl der Mandantenerklärung wurde bei Spezialisten somit nicht signifikant von deren Vererbung beeinflusst.

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Da die Mandantenerklärung im Rahmen der Nicht-Fehler-Version zugleich die tatsächliche Abweichungsursache ist, ist entsprechend zu konstatieren, dass von einer vererbten Hypothese auch ein sig-



|                                                         | Mit<br>Mandantenerklärung<br>n=10 | Ohne<br>Mandantenerklärung<br>n=11 | Signifikanz <sup>2</sup><br>(1-seitig) |
|---------------------------------------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 80,0                              | 63,6                               | 0,20                                   |

Tabelle 50: Auswahlhäufigkeit der Mandantenerklärung durch Spezialisten (Nicht-Fehler-Version)

#### 4.4.5.3 Test von $FF_{IV}$

##### 4.4.5.3.1 Fehler-Version

Tabelle 51 zeigt, dass Prüfer mit Mandantenerklärung diese häufiger auswählten, wenn ein geringes Prüfungsrisiko vorliegt (45,5% ggü. 36,4%;  $p = 0,33$ ).<sup>1</sup> Obwohl dieser Unterschied nicht signifikant ist, weist er zumindest in die von  $H_{IV.1.1}$  und  $H_{IV.1.2}$  vermutete Richtung. Weniger signifikant sind die Ergebnisse indes, wenn ein erhöhtes Prüfungsrisiko vorliegt (27,3% ggü. 30,0%; 0,46).<sup>2</sup> Demzufolge liegt auch hier die Vermutung nahe, dass ein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss des Erhalts einer Mandantenerklärung auf die Auswahl der Mandantenerklärung als wahrscheinlichste Abweichungsursache existiert. Tendenziell ist diese Beeinflussung stärker einzuschätzen, wenn ein geringes Prüfungsrisiko vorliegt. Ein erhöhtes Prüfungsrisiko könnte diesen Tendenzen indes entgegenwirken.

|                                                         | Mit<br>Mandantenerklärung |              | Ohne<br>Mandantenerklärung |              |
|---------------------------------------------------------|---------------------------|--------------|----------------------------|--------------|
|                                                         | Niedrig<br>n=11           | Hoch<br>n=11 | Niedrig<br>n=11            | Hoch<br>n=10 |
| <b>Auswahlhäufigkeit<br/>Mandantenerklärung<br/>[%]</b> | 45,5                      | 27,3         | 36,4                       | 30,0         |

Tabelle 51: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

nifikanter Einfluss auf die Erfolgsquote der Nicht-Spezialisten ausgeht. Siehe hierzu auch die Ausführungen unter Kap. 4.4.6.2.2.

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

#### 4.4.5.3.2 Nicht-Fehler-Version

Anhand von Tabelle 52 ist zu erkennen, dass die vererbte Hypothese bei erhöhtem Prüfungsrisiko signifikant häufiger von Prüfern mit Mandantenerklärung als von Prüfern ohne Mandantenerklärung selektiert wird (75% ggü. 41,7%;  $p = 0,05$ ).<sup>3</sup> Demgegenüber liegt kein signifikanter Einfluss des Erhalts einer Mandantenerklärung vor, wenn ein niedriges Prüfungsrisiko zu berücksichtigen ist (70% ggü. 50%;  $p = 0,18$ ).<sup>4</sup> Folglich kann auch hier die Vermutung aufgestellt werden, dass ein Zusammenhang zwischen dem Prüfungsrisiko und dem Einfluss des Erhalts einer Mandantenerklärung auf deren Selektion besteht. Allerdings weisen die Ergebnisse hier auf einen umgekehrten Ursache-Wirkungszusammenhang hin. Möglicherweise wird der Einfluss des Erhalts einer Mandantenerklärung durch erhöhtes Prüfungsrisiko intensiviert.

|                                                 | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|-------------------------------------------------|------------------------|--------------|-------------------------|--------------|
|                                                 | Niedrig<br>n=10        | Hoch<br>n=12 | Niedrig<br>n=10         | Hoch<br>n=12 |
| <b>Auswahlhäufigkeit Mandantenerklärung [%]</b> | 70,0                   | 75,0         | 50,0                    | 41,7         |

Tabelle 52: Auswahlhäufigkeit der Mandantenerklärung durch Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version)

#### 4.4.6 Erfolgsquote

##### 4.4.6.1 Test von $H_{V,I}$

##### 4.4.6.1.1 Fehler-Version

Wie aus Tabelle 53 hervorgeht, erreichen Prüfer mit Mandantenerklärung eine Erfolgsquote von 40,9% im Vergleich zu 47,6% bei Probanden ohne Mandantenerklärung. Obwohl die Ergebnisse in die von  $H_{V,I}$  vermutete Richtung gehen, sind die Unterschiede nicht signifikant ( $p = 0,33$ ).  $H_{V,I}$  ist folglich abzulehnen.

<sup>1</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>2</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>3</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>4</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

|                             | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=21 | Signifikanz <sup>1</sup><br>(1-seitig) |
|-----------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Erfolgsquote<br/>[%]</b> | 40,9                              | 47,6                               | 0,33                                   |

Tabelle 53: Erfolgsquote der Prüfer (Fehler-Version)

#### 4.4.6.1.2 Nicht-Fehler-Version

Aus Tabelle 54 geht hervor, dass 72,7% der Prüfer mit Mandantenerklärung die analytische Prüfungshandlung erfolgreich abschließen. Demgegenüber weisen Prüfer ohne Mandantenerklärung eine Erfolgsquote von lediglich 45,5% auf ( $p = 0,03$ ).  $H_{V,I}$  ist folglich anzunehmen.

|                             | Mit<br>Mandantenerklärung<br>n=22 | Ohne<br>Mandantenerklärung<br>n=22 | Signifikanz <sup>2</sup><br>(1-seitig) |
|-----------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Erfolgsquote<br/>[%]</b> | 72,7                              | 45,5                               | 0,03                                   |

Tabelle 54: Erfolgsquote der Prüfer (Nicht-Fehler-Version)

#### 4.4.6.2 Test von $H_{V,I,1}$ und $H_{V,I,2}$ <sup>3</sup>

##### 4.4.6.2.1 Fehler-Version

###### *Nicht-Spezialisten*

Entgegen der Erwartung von  $H_{V,I,1}$  wählten Nicht-Spezialisten mit Mandantenerklärung den tatsächlichen Abweichungsgrund genauso häufig aus wie Nicht-Spezialisten ohne Mandantenerklärung (33,3% ggü. 33,3%;  $p = 1,00$ ). Somit ist  $H_{V,I,1}$  zurückzuweisen. Ein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Erfolgsquote von Nicht-Spezialisten war folglich nicht evident.

---

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

<sup>3</sup> Da im Rahmen der Nicht-Fehler-Version eine Identität zwischen Mandantenerklärung und tatsächlichem Abweichungsgrund vorliegt, können  $H_{V,I,1}$  sowie  $H_{V,I,2}$  hier keiner Überprüfung unterzogen werden. Um dennoch Aussagen über die Erfolgsquote von Nicht-Spezialisten einerseits und Spezialis-

|                             | Mit<br>Mandantenerklärung<br>n=9 | Ohne<br>Mandantenerklärung<br>n=9 | Signifikanz <sup>1</sup><br>(1-seitig) |
|-----------------------------|----------------------------------|-----------------------------------|----------------------------------------|
| <b>Erfolgsquote<br/>[%]</b> | 33,3                             | 33,3                              | 1,00                                   |

Tabelle 55: Erfolgsquote der Nicht-Spezialisten (Fehler-Version)

### Spezialisten

Wie Tabelle 56 zeigt, selektierten Spezialisten, denen eine Mandantenerklärung zur Verfügung gestellt wurde, die tatsächliche Abweichungsursache seltener aus als Spezialisten ohne Mandantenerklärung (46,2% ggü. 53,8;  $p = 0,27$ ). Obwohl die Ergebnisse somit in die von  $H_{V.1.2}$  vermutete Richtung weisen, sind die Unterschiede jedoch nicht signifikant.  $H_{V.1.2}$  ist folglich abzulehnen.

|                             | Mit<br>Mandantenerklärung<br>n=13 | Ohne<br>Mandantenerklärung<br>n=12 | Signifikanz <sup>2</sup><br>(1-seitig) |
|-----------------------------|-----------------------------------|------------------------------------|----------------------------------------|
| <b>Erfolgsquote<br/>[%]</b> | 46,2                              | 53,8                               | 0,27                                   |

Tabelle 56: Erfolgsquote der Nicht-Spezialisten (Fehler-Version)

#### 4.4.6.2.2 Nicht-Fehler-Version

Da Mandantenerklärung und tatsächlicher Abweichungsgrund im Rahmen der Nicht-Fehler-Version identisch sind, können  $H_{V.1.1}$  und  $H_{V.1.2}$  hier nicht überprüft werden. Um dennoch eine Überprüfung des Zusammenhangs zwischen dem Erhalt einer Mandantenerklärung und der Erfolgsquote der Spezialisten und Nicht-Spezialisten durchführen zu können, soll an dieser Stelle auf  $H_{IV.1.1}$  und  $H_{IV.1.2}$  zurückgegriffen werden, da diese inhaltlich zugleich auf die Erfolgsquote gerichtet sind.

Wie aus Tabelle 49 hervorgeht, wiesen Nicht-Spezialisten mit Mandantenerklärung eine Erfolgsquote von 66,7% auf. Im Vergleich zu einer Erfolgsquote von 27,3% bei Nicht-Spezialisten ohne Mandantenerklärung ist dieser Unterschied signifikant ( $p = 0,03$ ). Folglich wurde  $H_{IV.1.1}$  bestätigt. Die Erfolgsquote von Nicht-Spezialisten wurde demnach signifikant von der Vererbung einer Mandantenerklärung beeinflusst.

---

ten andererseits treffen zu können, soll an dieser Stelle auf  $H_{IV.1.1}$  und  $H_{IV.1.2}$  zurückgegriffen, da sich die in ihnen zum Ausdruck kommende Vermutung inhaltlich auch auf die Erfolgsquote richtet.

<sup>1</sup> Chi-Quadrat-Test.

<sup>2</sup> Chi-Quadrat-Test.

Demgegenüber ist die Erfolgsquote von Spezialisten in Abhängigkeit vom Erhalt der Abweichungserklärung durch den Mandanten nicht signifikant unterschiedlich (80,0% ggü. 63,6%;  $p = 0,20$ ; siehe Tabelle 50). Obwohl die Ergebnisse in die vermutete Richtung weisen, ist  $H_{IV.1.2}$  abzulehnen. Signifikante Auswirkungen auf den Erfolg von Spezialisten waren demnach nicht zu verzeichnen.

#### 4.4.6.3 Test von $FF_V$

##### 4.4.6.3.1 Fehler-Version

Wie in Tabelle 57 zu erkennen ist, fallen die Unterschiede zwischen Prüfern mit und ohne Mandantenerklärung tendenziell stärker aus, wenn ein geringes Prüfungsrisiko vorliegt. Hier erreichen Prüfer mit Mandantenerklärung eine Erfolgsquote von 36,4% im Vergleich zu 45,5% bei Prüfern ohne Mandantenerklärung ( $p = 0,33$ ).<sup>1</sup> Liegt hingegen ein hohes Prüfungsrisiko vor, so sind die Unterschiede geringer (45,5% ggü. 50,0%;  $p = 0,44$ ).<sup>2</sup> Obwohl die Unterschiede jeweils nicht signifikant ausfallen, lassen die Ergebnisse die Existenz eines Zusammenhangs zwischen dem Einfluss des Erhalts einer Mandantenerklärung auf die Erfolgsquote und dem zugrunde liegenden Prüfungsrisiko vermuten. Dem effektivitätsmindernden Einfluss einer Mandantenerklärung im Rahmen des geringen Risikos kann zumindest zum Teil durch ein gesteigertes Risiko entgegengewirkt werden. Erhöhtem Prüfungsrisiko könnte möglicherweise ein debiasing-Mechanismus zugesprochen werden.

|                  | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|------------------|------------------------|--------------|-------------------------|--------------|
|                  | Niedrig<br>n=11        | Hoch<br>n=11 | Niedrig<br>n=11         | Hoch<br>n=10 |
| Erfolgsquote [%] | 36,4                   | 45,5         | 45,5                    | 50,0         |

Tabelle 57: Erfolgsquote der Prüfer unter Berücksichtigung des Prüfungsrisikos (Fehler-Version)

<sup>1</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

<sup>2</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

#### 4.4.6.3.2 Nicht-Fehler-Version

Wie in Tabelle 58 zu erkennen ist, fallen die Erfolgsunterschiede in Abhängigkeit des Erhalts einer Mandantenerklärung stärker aus, wenn ein erhöhtes Prüfungsrisiko vorliegt. Hier liegt die Erfolgsquote von Prüfern mit Mandantenerklärung bei 75% im Vergleich zu 41,7% bei Probanden ohne Mandantenerklärung ( $p = 0,05$ ).<sup>1</sup> Bei geringem Prüfungsrisiko erreichen Erstere eine Erfolgsquote von 70% und Letztere von 50% ( $p = 0,18$ ).<sup>1</sup> Diese Ergebnisse lassen folglich auf einen Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Erfolgsquote und dem zugrunde liegenden Prüfungsrisiko schließen. Obwohl die Erfolgsquote bereits bei geringem Risiko höher ausfällt, wenn dem Prüfer eine Mandantenerklärung zur Verfügung gestellt wird, legen die Befunde nahe, dass ein erhöhtes Prüfungsrisiko eine weitere Erhöhung der Effektivität bewirkt.

|                         | Mit Mandantenerklärung |              | Ohne Mandantenerklärung |              |
|-------------------------|------------------------|--------------|-------------------------|--------------|
|                         | Niedrig<br>n=10        | Hoch<br>n=12 | Niedrig<br>n=10         | Hoch<br>n=12 |
| <b>Erfolgsquote [%]</b> | 70,0                   | 75,0         | 50,0                    | 41,7         |

Tabelle 58: Erfolgsquote der Prüfer unter Berücksichtigung des Prüfungsrisikos (Nicht-Fehler-Version)

---

<sup>1</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

## **4.5 Zusammenfassung der Ergebnisse und Schlussfolgerungen**

### **4.5.1 Vorbemerkungen**

In der vorliegenden Studie konnte eine Vielzahl von Auswirkungen des Erhalts einer Nicht-Fehlererklärung durch den Mandanten auf die prüferische Vorgehensweise sowie auf die Effektivität und Effizienz der durchgeführten analytischen Prüfungshandlung identifiziert werden. Im Folgenden werden die Befunde getrennt nach den Phasen der (anfänglichen und gesamten) Hypothesengenerierung, der Informationssuche, der Hypothesenbewertung sowie der Erfolgsquote der Probanden dargestellt.

### **4.5.2 Anfängliche Hypothesengenerierung**

Im Rahmen der *anfänglichen Hypothesengenerierung* wurde festgestellt, dass der Erhalt einer Mandantenerklärung keinen signifikanten Einfluss auf die Anzahl der anfänglich generierten Nicht-Fehlerhypothesen hatte und somit kein spreading effect zu verzeichnen war. Dieses gilt unabhängig davon, ob die Branchenspezifität der Probanden berücksichtigt wurde oder nicht. Anders verhält es sich hingegen bei dem Einfluss einer Mandantenerklärung auf die Anzahl der anfänglich generierten Fehlerhypothesen. Ohne Berücksichtigung der branchenspezifischen Erfahrung konnte zunächst keine Auswirkung auf die Anzahl der generierten Fehlerhypothesen festgestellt werden. Differenziert man hingegen nach der Branchenspezifität der Probanden, konnte nachgewiesen werden, dass Spezialisten durch den Erhalt einer Mandantenerklärung eine signifikant höhere Anzahl von Fehlerhypothesen generierten (sog. „inverser“ spreading effect). Dieser Effekt konnte im Wesentlichen darauf zurückgeführt werden, dass Spezialisten das wahrgenommene Prüfungsrisiko wesentlich höher einschätzten, wenn sie unter Berücksichtigung eines erhöhten Prüfungsrisikos eine Nicht-Fehlererklärung des Mandanten erhielten. Möglicherweise führte das Vorliegen eines erhöhten Prüfungsrisikos und der gleichzeitige Versuch des Mandanten, die festgestellte Abweichung auf eine Nicht-Fehlerursache zurückzuführen, dazu, dass die Spezialisten eine gesteigerte professionelle Skepsis zeigten.

Ein anderes Bild ergibt sich hingegen, wenn nicht nach der Branchenspezifität der Probanden, sondern nach dem zugrunde liegenden Prüfungsrisiko differenziert wird. Hier wurde gezeigt, dass Prüfer unter Berücksichtigung eines geringen Prüfungsrisikos we-

---

<sup>1</sup> Chi-Quadrat-Test; 1-seitige Signifikanz.

sentlich mehr Nicht-Fehlerhypothesen generierten, wenn ihnen eine Mandantenerklärung mitgeteilt wurde. Dieser sog. spreading effect konnte bei erhöhten Prüfungsrisiko nicht nachgewiesen werden. Bei erhöhtem Prüfungsrisiko war hingegen evident, dass Prüfer mit Mandantenerklärung signifikant mehr Fehlerhypothesen generierten als Prüfer ohne Mandantenerklärung. Dieser sog. „inverse“ spreading effect konnte bei geringem Prüfungsrisiko indes nicht festgestellt werden.

Insgesamt ist zu konstatieren, dass der interference effect im Rahmen der vorliegenden Studie nicht nachgewiesen werden konnte. Vielmehr nehmen Spezialisten bzw. Prüfer den Erhalt einer Nicht-Fehlererklärung des Mandanten unter Berücksichtigung eines erhöhten Prüfungsrisikos zum Anlass, um eine erhöhte professionelle Skepsis walten zu lassen, die dazu führt, dass wesentlich mehr Fehlerhypothesen generiert werden. Insofern kann die Befürchtung, dass sich der Erhalt einer Mandantenerklärung negativ auswirkt nur bedingt geteilt werden.

In Bezug auf den spreading effect weisen die Ergebnisse darauf hin, dass dieser ausschließlich bei einem geringen Prüfungsrisikos verzeichnet werden konnten. Der Erhalt einer Mandantenerklärung hatte hingegen keinen signifikanten Einfluss auf die Anzahl der Nicht-Fehlerhypothesen, wenn ein erhöhtes Prüfungsrisiko zu berücksichtigen war. Insofern kann die Befürchtung, dass sich der Erhalt einer Mandantenerklärung negativ auswirkt auch nur bedingt geteilt werden.

In der nachfolgenden Übersicht werden die Ergebnisse zur anfänglichen Hypothesengenerierung unter Berücksichtigung der angewandten Testverfahren und der entsprechenden Signifikanzniveaus tabellarisch wiedergegeben:



| Hypothese/Forschungsfrage                                                                                                                                                                                                   | Bez.        | Testverfahren       | Ergebnis | Erläuterung                                                                                                                                                                                                                                                                      | Sign.        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                            | $H_{I,1}$   | Mann-Whitney-U-Test | –        | Kein interference effect                                                                                                                                                                                                                                                         | 0,32         |
| Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                         | $H_{I,2}$   | Mann-Whitney-U-Test | –        | Kein spreading effect                                                                                                                                                                                                                                                            | 0,27         |
| Nicht-Spezialisten, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.    | $H_{I,1,1}$ | Mann-Whitney-U-Test | –        | Kein interference effect                                                                                                                                                                                                                                                         | 0,20         |
| Nicht-Spezialisten, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten. | $H_{I,2,1}$ | Mann-Whitney-U-Test | –        | Kein spreading effect                                                                                                                                                                                                                                                            | 0,42         |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich weniger Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                         | $H_{I,1,2}$ | Mann-Whitney-U-Test | –        | Kein interference effect<br>(sog. „inverser“ spreading effect insbesondere bei Spezialisten mit Mandantenerklärung und erhöhtem Prüfungsrisiko, da hier das wahrgenommene Prüfungsrisiko höher ausfiel als bei Spezialisten ohne Mandantenerklärung und erhöhtem Prüfungsrisiko) | 0,08         |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren anfänglich mehr Nicht-Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                      | $H_{I,2,2}$ | Mann-Whitney-U-Test | –        | Kein spreading effect                                                                                                                                                                                                                                                            | 0,28         |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die anfängliche Hypothesengenerierung und dem Prüfungsrisiko?                                                                                   | $FF_{I,1}$  | Mann-Whitney-U-Test | n.a.     | Die anfängliche Hypothesengenerierung ist nur bei geringem Prüfungsrisiko vom spreading effect betroffen<br>Die anfängliche Hypothesengenerierung ist nur bei erhöhtem Prüfungsrisiko vom „inversen“ spreading effect betroffen                                                  | 0,10<br>0,09 |

Tabelle 59: Zusammenfassung der Untersuchungsergebnisse (anfängliche Hypothesengenerierung)

### 4.5.3 Gesamte Hypothesengenerierung

In Bezug auf die *gesamte Hypothesengenerierung* ist festzuhalten, dass im Rahmen der Fehler-Version weder ein interference effect noch ein spreading effect nachgewiesen werden konnte. Dieser Befund war sowohl ohne Berücksichtigung der Branchenspezifität als auch mit Berücksichtigung der Branchenspezifität zu verzeichnen. Insofern sind die Befunde zur gesamten Hypothesengenerierung identisch zu den Befunden bei der anfänglichen Hypothesengenerierung. Ein differenziertes Bild ergibt sich hingegen unter Berücksichtigung des zugrunde liegenden Prüfungsrisikos. Während die Prüfer mit Mandantenerklärung im Rahmen der anfänglichen Hypothesengenerierung (noch) zu einem vermehrten Abruf von Nicht-Fehlerhypothesen neigten, wenn ein geringes Prüfungsrisiko vorlag, konnte dieser Effekt im Rahmen der gesamten Hypothesengenerierung nicht mehr nachgewiesen werden. Die Prüfer scheinen somit ihre anfängliche Beeinflussung durch die Mandantenerklärung kompensiert zu haben. Folglich ist für die Fehler-Version festzuhalten, dass der Erhalt einer Nicht-Fehlererklärung durch den Mandanten keine negativen Auswirkungen auf die gesamte Hypothesengenerierung mit sich brachte.

Im Rahmen der Nicht-Fehler-Versionen konnte hingegen festgestellt werden, dass der Erhalt einer Mandantenerklärung zu einer vermehrten Generierung von Nicht-Fehlerhypothesen führte und die gesamte Hypothesengenerierung im Gegensatz zur anfänglichen Hypothesengenerierung somit durch einen spreading effect gekennzeichnet war. Diese Beobachtung konnte im Wesentlichen auf das Verhalten der Spezialisten zurückgeführt werden. Nicht-Spezialisten ließen sich demgegenüber nicht durch den Erhalt einer Mandantenerklärung beeinflussen. Weiterhin konnte festgestellt werden, dass der spreading effect sowohl bei geringem Prüfungsrisiko als auch bei erhöhtem Prüfungsrisiko nachgewiesen werden konnte. Fraglich ist allerdings, wie dieser Effekt zu bewerten ist. Da die festgestellte Abweichung hier auf einen Nicht-Fehler zurückzuführen ist, kann davon ausgegangen werden, dass der vermehrte Abruf von Nicht-Fehlerhypothesen tendenziell zu einer höheren Erfolgswahrscheinlichkeit führt. Insofern ist der spreading effect hier als positiv zu bewerten. Ein interference effect konnte im Rahmen der Nicht-Fehler-Version indes nicht festgestellt werden.

Insgesamt ist demnach festzuhalten, dass der Erhalt einer Mandantenerklärung keine negativen Auswirkungen auf die gesamte Hypothesengenerierung bewirkte. Im Rahmen der Fehler-Version konnte weder ein interference effect noch ein spreading effect verzeichnet werden. Innerhalb der Nicht-Fehler-Version führte die Mandantenerklärung

sogar zu einer positiven Beeinflussung, da verstärkt Nicht-Fehlerhypothesen generiert wurden.

Tabelle 60 gibt eine Zusammenfassung der Ergebnisse zur gesamten Hypothesengenerierung unter Berücksichtigung der angewandten Testverfahren und der entsprechenden Signifikanzniveaus wieder:

| Hypothese/Forschungsfrage                                                                                                                                                                                         | Bez.        | Testverfahren       | Version      | Ergebnis | Erläuterung                                                                                                                                                                                                  | Sign.                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------------|--------------|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                   | $H_{I,3}$   | Mann-Whitney-U-Test | Fehler       | –        | Kein interference effect                                                                                                                                                                                     | 0,38                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | –        | Kein interference effect                                                                                                                                                                                     | 0,39                     |
| Prüfer, die zunächst eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Prüfer, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                | $H_{I,4}$   | Mann-Whitney-U-Test | Fehler       | –        | Kein spreading effect                                                                                                                                                                                        | 0,30                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | +        | Spreading effect                                                                                                                                                                                             | 0,046                    |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.    | $H_{I,3,1}$ | Mann-Whitney-U-Test | Fehler       | –        | Kein interference effect                                                                                                                                                                                     | 0,46                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | –        | Kein interference effect                                                                                                                                                                                     | 0,24                     |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Nicht-Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten. | $H_{I,4,1}$ | Mann-Whitney-U-Test | Fehler       | –        | Kein spreading effect                                                                                                                                                                                        | 0,43                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | –        | Kein spreading effect                                                                                                                                                                                        | 0,11                     |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt weniger Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.                | $H_{I,3,2}$ | Mann-Whitney-U-Test | Fehler       | –        | Kein interference effect                                                                                                                                                                                     | 0,41                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | –        | Kein interference effect                                                                                                                                                                                     | 0,15                     |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, generieren insgesamt mehr Nicht-Fehlerhypothesen als Spezialisten, die keine Nicht-Fehlererklärung durch den Mandanten erhalten.             | $H_{I,4,2}$ | Mann-Whitney-U-Test | Fehler       | –        | Kein spreading effect                                                                                                                                                                                        | 0,16                     |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | +        | Spreading effect                                                                                                                                                                                             | 0,10                     |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die gesamte Hypothesengenerierung und dem Prüfungsrisiko?                                                                             | $FF_{I,2}$  | Mann-Whitney-U-Test | Fehler       | n.a.     | Interference effect oder spreading effect konnten weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden                                                                                   | 0,41/0,27 bzw. 0,43/0,19 |
|                                                                                                                                                                                                                   |             |                     | Nicht-Fehler | n.a.     | Interference effect konnte weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden<br>Spreading effect konnte sowohl bei niedrigem als auch bei erhöhtem Prüfungsrisiko nachgewiesen werden | 0,31/0,22<br>0,10/0,09   |

Tabelle 60: Zusammenfassung der Untersuchungsergebnisse (gesamte Hypothesengenerierung)

#### 4.5.4 Informationssuche

##### 4.5.4.1 Angewandte Suchstrategie

In der vorliegenden Studie konnte bei Prüfern (d.h. ohne Berücksichtigung der Branchenspezifität und ohne Berücksichtigung des Prüfungsrisikos) keine wesentlichen Einflüsse des Erhalts einer Mandantenerklärung auf die angewandte Suchstrategie nachgewiesen werden. Insbesondere führte die Mandantenerklärung weder zu einer verstärkten Suche nach Informationen, die auf die Beurteilung der Mandantenerklärung gerichtet waren, noch zu einer eingeschränkten Suche nach Informationen, die auf andere potenzielle Abweichungsursachen gerichtet waren. Wurde hingegen nach der Branchenspezifität der Probanden differenziert, so konnte festgestellt werden, dass Spezialisten mit Mandantenerklärung im Rahmen der *Fehler-Version* mehr Tests zur Beurteilung der Mandantenerklärung durchführten. Da hiervon jedoch weder ein signifikanter Einfluss auf die Anzahl der durchgeführten Tests noch auf die in Anspruch genommene Prüfzeit ausging, ist davon auszugehen, dass die verstärkte Suche nach Informationen zur Beurteilung der Mandantenerklärung keine negativen Auswirkungen auf die Effizienz der Spezialisten mit sich brachte.

Eine andere Beurteilung ist hingegen bei der Vorgehensweise der Spezialisten im Rahmen der *Nicht-Fehler-Version* vorzunehmen. Zwar konnte hier kein Einfluss der Mandantenerklärung auf die Suche nach Informationen zur Beurteilung der Mandantenerklärung einerseits und die Suche nach Informationen zur Beurteilung anderer potenzieller Abweichungsursachen andererseits verzeichnet werden. Allerdings führte der Erhalt einer Mandantenerklärung bei Spezialisten zu einer signifikant höheren Anzahl der durchgeführten Kat. 4-Tests (Nicht-Fehler) und Kat. 5-Test (Fehler). Die vermehrte Untersuchung von Fehlerhypothesen könnte auf den im Rahmen der anfänglichen Hypothesengenerierung identifizierten „inversen“ spreading effect zurückgeführt werden. Die anschließend verstärkte Untersuchung von Nicht-Fehlerhypothesen könnte durch den im Rahmen der gesamten Hypothesengenerierung festgestellten spreading effect hervorgerufen worden sein. Diese verstärkte Tiefensuche der Spezialisten führte schließlich zu einer signifikant höheren Anzahl der durchgeführten Tests insgesamt und der in Anspruch genommenen Prüfzeit. Insofern weisen die Ergebnisse darauf hin, dass der Erhalt einer Mandantenerklärung innerhalb der *Nicht-Fehler-Version* einen negativen Einfluss auf die Effizienz von analytischen Prüfungshandlungen haben könnte.

Differenziert man hingegen nicht nach der Branchenspezifität der Prüfer, sondern nach dem zugrunde liegenden Prüfungsrisiko, so ist festzustellen, dass im Rahmen der Feh-

ler-Version keine signifikanten Einflüsse einer Mandantenerklärung auf die Suche nach Informationen zur Beurteilung der Mandantenerklärung einerseits und zur Beurteilung anderer potenzieller Abweichungsursachen andererseits nachgewiesen werden konnten. Im Rahmen der Nicht-Fehler-Version bewirkte der Erhalt einer Mandantenerklärung hingegen eine verstärkte Einholung von Prüfungsnachweisen zur Beurteilung der Mandantenerklärung, wenn ein erhöhtes Prüfungsrisiko zu berücksichtigen war. Diese auf die Mandantenerklärung gerichtete Suchstrategie konnte bei geringem Prüfungsrisiko hingegen nicht festgestellt werden. Möglicherweise führt der Erhalt einer Nicht-Fehlererklärung durch den Mandanten bei erhöhtem Prüfungsrisiko dazu, dass Prüfer mehr Prüfungsnachweise einholen wollen, die die Korrektheit der Mandantenerklärung bestätigen. Liegt hingegen ein geringes Prüfungsrisiko vor, reichen den Prüfern weniger Prüfungsnachweise, um die Angemessenheit der Mandantenerklärung zu akzeptieren. Da die positive Suchstrategie bei erhöhtem Risiko zu einer signifikant höheren Prüfzeit führte, ist auch hier davon auszugehen, dass der Erhalt einer Mandantenerklärung eine mögliche Beeinträchtigung der Effizienz von analytischen Prüfungshandlungen mit sich bringt.

Eine Zusammenfassung der Ergebnisse zur angewandten Suchstrategie unter Berücksichtigung der Testverfahren und der entsprechenden Signifikanzniveaus findet sich in folgender Übersicht:

| Hypothese/Forschungsfrage                                                                                                                                                                                                                 | Bez.         | Testverfahren       | Version      | Ergebnis | Erläuterung                                                                                                                                                                                                                                                  | Sign.                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|---------------------|--------------|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Prüfer ohne Mandantenerklärung.                                           | $H_{II,1}$   | Mann-Whitney-U-Test | Fehler       | –        | Keine verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                          | 0,21                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                          | 0,16                   |
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Prüfer ohne Mandantenerklärung.                         | $H_{II,2}$   | Mann-Whitney-U-Test | Fehler       | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,57                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,37                   |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Nicht-Spezialisten ohne Mandantenerklärung.                   | $H_{II,1,1}$ | Mann-Whitney-U-Test | Fehler       | –        | Keine verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                          | 0,42                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                          | 0,11                   |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Nicht-Spezialisten ohne Mandantenerklärung. | $H_{II,2,1}$ | Mann-Whitney-U-Test | Fehler       | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,48                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,21                   |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen mehr Informationen ein, die auf die Beurteilung der Mandantenerklärung gerichtet sind, als Spezialisten ohne Mandantenerklärung.                               | $H_{II,1,2}$ | Mann-Whitney-U-Test | Fehler       | +        | Verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                                | 0,09                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung                                                                                                                                                                                          | 1,00                   |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, holen weniger Informationen ein, die auf die Beurteilung anderer potenzieller Fehlerhypothesen gerichtet sind, als Spezialisten ohne Mandantenerklärung              | $H_{II,2,2}$ | Mann-Whitney-U-Test | Fehler       | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,24                   |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | –        | Keine eingeschränkte Breitensuche                                                                                                                                                                                                                            | 0,11                   |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die angewandte Suchstrategie und dem Prüfungsrisiko?                                                                                                          | $FF_{II}$    | Mann-Whitney-U-Test | Fehler       | n.a.     | Verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung konnte weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden<br>Eingeschränkte Breitensuche konnte weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden | 0,12/1,00<br>0,73/0,68 |
|                                                                                                                                                                                                                                           |              |                     | Nicht-Fehler | n.a.     | Verstärkte Tiefensuche zur Beurteilung der Mandantenerklärung konnte lediglich bei erhöhtem Prüfungsrisiko nachgewiesen werden<br>Eingeschränkte Breitensuche konnte weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden                | 0,05/1,00<br>0,22/0,41 |

Tabelle 61: Zusammenfassung der Untersuchungsergebnisse (angewandte Suchstrategie)

#### 4.5.4.2 Interpretation von Prüfungsnachweisen

Im Rahmen der Interpretation von Prüfungsnachweisen konnte ein durch den Erhalt einer Mandantenerklärung verursachter confirmation bias ausschließlich im Rahmen der Nicht-Fehler-Version nachgewiesen werden. Dieses ist im Wesentlichen auf das Verhalten der Spezialisten zurückzuführen. Allerdings ist auch hier erneut darauf hinzuweisen, dass dieser Effekt nicht als negativ bewertet werden kann, da die Mandantenerklärung zugleich der tatsächliche Abweichungsgrund ist. In diesem Fall ist der Erhalt einer (korrekten) Mandantenerklärung vielmehr als positiv zu bezeichnen. Spezialisten mit Mandantenerklärung scheinen bei entsprechender Informationslage besser dazu in der Lage zu sein, die Korrektheit einer Mandantenerklärung zu erkennen als Spezialisten ohne Mandantenerklärung.

Unterscheidet man hingegen nach dem zugrunde liegenden Prüfungsrisiko, so ist festzustellen, dass ein unterschiedliches Risiko lediglich im Rahmen der Nicht-Fehler-Version zu verschiedenen Ergebnissen führt. Hier kann ein confirmation bias ausschließlich unter Berücksichtigung eines geringen Prüfungsrisikos verzeichnet werden. Bei erhöhtem Risiko war dieser Effekt hingegen nicht mehr evident. Möglicherweise hat das erhöhte Prüfungsrisiko dazu geführt, dass die Prüfer den zur Beurteilung der Mandantenerklärung eingeholten Prüfungsnachweisen skeptischer gegenüber stehen als bei geringem Prüfungsrisiko. Da es sich bei der Mandantenerklärung jedoch um die tatsächliche Abweichungsursache handelte, ist die erhöhte professionelle Skepsis nicht angebracht. Vielmehr führt die kritische Haltung der Prüfer zu einer häufigeren Fehlinterpretation von Prüfungsnachweisen, die auf die richtige Abweichungsursache hindeuten.

Das Phänomen des „explaining away“ disconfirming evidence, falls eine Mandantenerklärung vererbt wurde, konnte hingegen weder unter Berücksichtigung der Branchenspezifität noch unter Berücksichtigung des Prüfungsrisikos festgestellt werden. Vielmehr ließen die Ergebnisse erkennen, dass Prüfer mit Mandantenerklärung unter Berücksichtigung eines erhöhten Prüfungsrisikos Informationen, die auf das Vorliegen einer anderen als der durch den Mandanten gelieferten Abweichungsursache hindeuten, häufiger richtig interpretierten als Prüfer ohne Mandantenerklärung.

In der nachfolgenden Übersicht werden die Ergebnisse zur Interpretation von Prüfungsergebnissen unter Berücksichtigung der angewandten Testverfahren und der entsprechenden Signifikanzniveaus tabellarisch wiedergegeben:



| Hypothese/Forschungsfrage                                                                                                                                                                                                                                                  | Bez.          | Testverfahren    | Version      | Ergebnis | Erläuterung                                                                                                                                                                                                                                                                                                                                                                                                                                 | Sign.                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|------------------|--------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Prüfer ohne Mandantenerklärung.                                       | $H_{III,1}$   | Chi-Quadrat-Test | Fehler       | –        | Keine konfirmatorischen Prozesse                                                                                                                                                                                                                                                                                                                                                                                                            | 0,23                   |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | +        | Konfirmatorische Prozesse                                                                                                                                                                                                                                                                                                                                                                                                                   | 0,06                   |
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Prüfer ohne Mandantenerklärung.                         | $H_{III,2}$   | Chi-Quadrat-Test | Fehler       | –        | Kein sog. „explaining away“ disconfirming evidence                                                                                                                                                                                                                                                                                                                                                                                          | 0,28                   |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | n.a.     | n.a.                                                                                                                                                                                                                                                                                                                                                                                                                                        | n.a.                   |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Nicht-Spezialisten ohne Mandantenerklärung.               | $H_{III,1,1}$ | Chi-Quadrat-Test | Fehler       | –        | Keine konfirmatorischen Prozesse                                                                                                                                                                                                                                                                                                                                                                                                            | 0,15                   |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | –        | Keine konfirmatorischen Prozesse                                                                                                                                                                                                                                                                                                                                                                                                            | 0,25                   |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Nicht-Spezialisten ohne Mandantenerklärung. | $H_{III,2,1}$ | Chi-Quadrat-Test | Fehler       | –        | Kein sog. „explaining away“ disconfirming evidence                                                                                                                                                                                                                                                                                                                                                                                          | 0,04 <sup>1</sup>      |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | n.a.     | n.a.                                                                                                                                                                                                                                                                                                                                                                                                                                        | n.a.                   |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung dieser Nicht-Fehlererklärung gerichtete Information häufiger als bestätigende Information als Spezialisten ohne Mandantenerklärung.                           | $H_{III,1,2}$ | Chi-Quadrat-Test | Fehler       | –        | Keine konfirmatorischen Prozesse                                                                                                                                                                                                                                                                                                                                                                                                            | 0,30                   |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | +        | Konfirmatorische Prozesse                                                                                                                                                                                                                                                                                                                                                                                                                   | 0,08                   |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, interpretieren eine auf die Beurteilung der tatsächlichen Abweichungsursache gerichtete Information häufiger als nicht bestätigende Information als Spezialisten ohne Mandantenerklärung.             | $H_{III,2,2}$ | Chi-Quadrat-Test | Fehler       | –        | Kein sog. „explaining away“ disconfirming evidence                                                                                                                                                                                                                                                                                                                                                                                          | 0,23                   |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | n.a.     | n.a.                                                                                                                                                                                                                                                                                                                                                                                                                                        | n.a.                   |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Interpretation von Informationen und dem Prüfungsrisiko?                                                                                                                                   | $FF_{III}$    | Chi-Quadrat-Test | Fehler       | n.a.     | Konfirmatorische Prozesse konnten weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden<br>“explaining away“ disconfirming evidence konnte weder bei niedrigem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden. Vielmehr werden bei erhöhtem Prüfungsrisiko Informationen, die auf das Vorliegen einer anderen als der durch den Mandanten gelieferten Abweichungsursache hindeuten, häufiger richtig interpretiert | 0,34/0,28<br>0,09/0,27 |
|                                                                                                                                                                                                                                                                            |               |                  | Nicht-Fehler | n.a.     | Konfirmatorische Prozesse konnten lediglich bei geringem Prüfungsrisiko nachgewiesen werden                                                                                                                                                                                                                                                                                                                                                 | 0,09/0,24              |

Tabelle 62: Zusammenfassung der Untersuchungsergebnisse (Interpretation der Prüfungsergebnisse)

<sup>1</sup> Allerdings weist dieses Ergebnis in die entgegengesetzte Richtung. Daher ist diese Hypothese abzulehnen.

#### **4.5.5 Hypothesenbewertung**

In Bezug auf die Hypothesenbewertung konnte festgestellt werden, dass ein signifikanter Einfluss des Erhalts einer Nicht-Fehlererklärung durch den Mandanten ausschließlich im Rahmen der Nicht-Fehler-Version evident war. Hier wählten Prüfer mit Mandantenerklärung wesentlich häufiger diese Abweichungshypothese als wahrscheinlichste Ursache für die festgestellte Abweichung aus als Prüfer ohne Mandantenerklärung. Diese Beobachtung konnte im Wesentlichen auf das Verhalten der Nicht-Spezialisten zurückgeführt werden. Da die Mandantenerklärung jedoch die tatsächliche Abweichungsursache ist, kann hier nicht die Befürchtung geäußert werden, dass der Erhalt einer (korrekten) Mandantenerklärung einen negativen Einfluss auf die Hypothesenbewertung ausübt. Spezialisten ließen sich hingegen weder in der Fehler- noch in der Nicht-Fehler-Version durch den Erhalt einer Mandantenerklärung beeinflussen. Insofern weisen die Ergebnisse der vorliegenden Studie darauf hin, dass der Erhalt einer Mandantenerklärung keine negativen Auswirkungen auf die Hypothesenbewertung mit sich brachte.

Differenziert man nach dem zugrunde liegenden Prüfungsrisiko, so ist festzuhalten, dass ein signifikanter Einfluss ausschließlich im Rahmen der Nicht-Fehler-Version festgestellt werden konnte. Hier wählten Prüfer die Mandantenerklärung wesentlich häufiger als Abweichungsgrund aus, wenn ein erhöhtes Prüfungsrisiko zu berücksichtigen war.

Tabelle 63 gibt eine Zusammenfassung der Ergebnisse zur Hypothesenbewertung unter Berücksichtigung der angewandten Testverfahren und der entsprechenden Signifikanzniveaus wieder:

| Hypothese/Forschungsfrage                                                                                                                                                                               | Bez.         | Testverfahren    | Version      | Ergebnis | Erläuterung                                                                                                                                    | Sign.     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|------------------|--------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Prüfer ohne Mandantenerklärung.                         | $H_{IV.1}$   | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                     | 0,42      |
|                                                                                                                                                                                                         |              |                  | Nicht-Fehler | +        | Signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                          | 0,03      |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Nicht-Spezialisten ohne Mandantenerklärung. | $H_{IV.1.1}$ | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                     | 1,00      |
|                                                                                                                                                                                                         |              |                  | Nicht-Fehler | +        | Signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                          | 0,03      |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen diese Erklärung häufiger als wahrscheinlichste Abweichungsursache aus als Spezialisten ohne Mandantenerklärung.             | $H_{IV.1.2}$ | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                     | 0,37      |
|                                                                                                                                                                                                         |              |                  | Nicht-Fehler | –        | Kein signifikanter Einfluss des Erhalts einer Mandantenerklärung auf die Auswahlhäufigkeit                                                     | 0,20      |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die anschließende Bewertung dieser Erklärung und dem Prüfungsrisiko?                                                        | $FF_{IV}$    | Chi-Quadrat-Test | Fehler       | n.a.     | Ein signifikanter Einfluss des Erhalts einer Mandantenerklärung konnte weder bei geringem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden | 0,33/0,46 |
|                                                                                                                                                                                                         |              |                  | Nicht-Fehler | n.a.     | Ein signifikanter Einfluss des Erhalts einer Mandantenerklärung konnte lediglich bei erhöhtem Prüfungsrisiko nachgewiesen werden               | 0,05/0,18 |

Tabelle 63: Zusammenfassung der Untersuchungsergebnisse (Hypothesenbewertung)

#### **4.5.6 Erfolgsquote**

In Bezug auf die Erfolgsquote und somit die Effektivität der Prüfer weisen die Ergebnisse darauf hin, dass der Erhalt einer Mandantenerklärung lediglich im Rahmen der Nicht-Fehler-Version einen signifikanten Einfluss auf das Verhalten der Prüfer bewirkte. Hier selektierten Prüfer die korrekte Abweichungshypothese wesentlich häufiger, wenn ihnen diese durch den Mandanten zur Verfügung gestellt wurde. Diese Beobachtung konnte im Wesentlichen auf das Verhalten der Nicht-Spezialisten zurückgeführt werden. Möglicherweise versetzte der Erhalt der korrekten Abweichungsursache durch den Mandanten die Nicht-Spezialisten erst in die Lage, den tatsächlichen Grund einzugrenzen und schließlich zu identifizieren. Insofern ist zu konstatieren, dass der Erhalt einer (korrekten) Mandantenerklärung zu einer Erhöhung der Effektivität der Nicht-Spezialisten führte. Da gleichzeitig kein negativer Einfluss vom Erhalt einer Mandantenerklärung auf die Erfolgsquote der Spezialisten nachgewiesen werden konnte, ist insgesamt zu konstatieren, dass vom Erhalt einer Mandantenerklärung nicht nur kein negativer Einfluss auf die Effektivität von analytischen Prüfungshandlungen ausgeht. Vielmehr wird die Effektivität durch den Erhalt einer Mandantenerklärung erhöht und zwar dann, wenn die Erklärung des Mandanten korrekt ist.

Differenziert man nach dem zugrunde liegenden Prüfungsrisiko, so ist festzuhalten, dass ein signifikanter Einfluss ausschließlich im Rahmen der Nicht-Fehler-Version festgestellt werden konnte. Hier wählten Prüfer mit Mandantenerklärung die korrekte Abweichungsursache wesentlich häufiger aus als Prüfer ohne Mandantenerklärung, wenn ein erhöhtes Prüfungsrisiko zu berücksichtigen war.

Eine Zusammenfassung der Ergebnisse zur Erfolgsquote unter Berücksichtigung der Testverfahren und der entsprechenden Signifikanzniveaus findet sich in folgender Übersicht:

| Hypothese/Forschungsfrage                                                                                                                                                                                       | Bez.        | Testverfahren    | Version      | Ergebnis | Erläuterung                                                                                                                                    | Sign.     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|------------------|--------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Prüfer, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus als Prüfer ohne Mandantenerklärung.                         | $H_{V,1}$   | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss                                                                                                                    | 0,33      |
|                                                                                                                                                                                                                 |             |                  | Nicht-Fehler | +        | Signifikanter Einfluss                                                                                                                         | 0,03      |
| Nicht-Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus als Nicht-Spezialisten ohne Mandantenerklärung. | $H_{V,1.1}$ | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss                                                                                                                    | 1,00      |
|                                                                                                                                                                                                                 |             |                  | Nicht-Fehler | +        | Signifikanter Einfluss                                                                                                                         | 0,03      |
| Spezialisten, die eine Nicht-Fehlererklärung des Mandanten erhalten, wählen seltener den tatsächlichen Grund als wahrscheinlichste Abweichungsursache aus als Spezialisten ohne Mandantenerklärung.             | $H_{V,1.2}$ | Chi-Quadrat-Test | Fehler       | –        | Kein signifikanter Einfluss                                                                                                                    | 0,27      |
|                                                                                                                                                                                                                 |             |                  | Nicht-Fehler | –        | Kein signifikanter Einfluss                                                                                                                    | 0,20      |
| Besteht ein Zusammenhang zwischen dem Einfluss einer Mandantenerklärung auf die Erfolgsquote und dem Prüfungsrisiko?                                                                                            | $FF_V$      | Chi-Quadrat-Test | Fehler       | n.a.     | Ein signifikanter Einfluss des Erhalts einer Mandantenerklärung konnte weder bei geringem noch bei erhöhtem Prüfungsrisiko nachgewiesen werden | 0,33/0,44 |
|                                                                                                                                                                                                                 |             |                  | Nicht-Fehler | n.a.     | Ein signifikanter Einfluss des Erhalts einer Mandantenerklärung konnte lediglich bei erhöhtem Prüfungsrisiko nachgewiesen werden               | 0,05/0,18 |

Tabelle 64: Zusammenfassung der Untersuchungsergebnisse (Erfolgsquote)

## 4.6 Diskussion

In Anbetracht der vorliegenden Ergebnisse ist die Zweckmäßigkeit der durch IDW PS 312.27 und ISA 520.18 empfohlenen Mandantenbefragung differenziert zu betrachten. In Bezug auf die Effektivität von analytischen Prüfungshandlungen konnte kein negativer Einfluss des Erhalts einer Mandantenerklärung nachgewiesen werden. Vielmehr konnte die Effektivität durch den Erhalt einer Mandantenerklärung erhöht werden, wenn es sich dabei um den tatsächlichen Grund der festgestellten Abweichung handelte. Dieses ist im Wesentlichen auf das Verhalten der Nicht-Spezialisten zurückzuführen, die erst durch den Erhalt der korrekten Abweichungsursache durch den Mandanten in der Lage waren, die analytische Prüfungshandlung erfolgreich abzuschließen. Die Effektivität der Spezialisten wurde hingegen weder in der Fehler- noch in der Nicht-Fehler-Version wesentlich beeinträchtigt.

Hinsichtlich der Effizienz von analytischen Prüfungshandlungen hatte der Erhalt einer Mandantenerklärung lediglich im Rahmen der Nicht-Fehler-Version einen negativen Einfluss, was im Wesentlichen auf das Verhalten der Spezialisten zurückzuführen ist. Sie führten wesentlich mehr Prüfungstests durch und nahmen dabei wesentlich mehr Prüfungszeit in Anspruch, wenn ihnen eine Mandantenerklärung bereitgestellt wurde. Die Erklärung für dieses Verhalten liegt in der erhöhten subjektiven Risikowahrnehmung von Spezialisten, wenn der Mandant die festgestellte Abweichung auf eine Nicht-Fehlerursache zurückführt. Hierdurch wird eine verstärkte Generierung von Fehlerhypothesen hervorgerufen („inverser“ spreading effect), die eine umfangreiche Suche nach Informationen zu ihrer Bestätigung oder Widerlegung bedingen. Da die Abweichung jedoch durch einen Nicht-Fehler verursacht wurde, werden ihre Vermutungen nicht bestätigt. Vielmehr erlangen die Spezialisten ausschließlich Prüfungsnachweise, die auf die Nicht-Existenz eines Fehlers in der Rechnungslegung des Mandanten hindeuten. Daraufhin kehrt sich die erhöhte professionelle Skepsis der Spezialisten um, was sich darin äußert, dass anschließend mehr Nicht-Fehlerhypothesen generiert werden (spreading effect), die wiederum eine zusätzliche Informationssuche bewirken. Insgesamt führt der Erhalt einer korrekten Abweichungsursache durch den Mandanten zu einem umfangreicheren Suchverhalten der Spezialisten, welches die Effizienz der analytischen Prüfungshandlungen gefährdet.

Unter Berücksichtigung des Prüfungsrisikos ist die Zweckmäßigkeit der Mandantenbefragung ebenfalls differenziert zu betrachten. Innerhalb der Fehler-Version konnten keine negativen Auswirkungen auf die Effektivität und Effizienz der analytischen Prü-

fungshandlung nachgewiesen werden. Im Rahmen der Nicht-Fehler-Version führte der Erhalt der Mandantenerklärung bei erhöhtem Risiko hingegen zu einer vermehrten Suche nach Informationen zur Beurteilung der Mandantenerklärung. Hierdurch wurde die Anzahl der durchgeführten Tests und der Umfang der Prüfzeit wesentlich erhöht. Dieser Zeit- und somit Kostenerhöhung steht jedoch eine höhere Effektivität von Prüfern gegenüber, die eine Mandantenerklärung bei erhöhtem Prüfungsrisiko erhielten. Insofern ist hier ein trade-off zwischen der Effektivität von analytischen Prüfungshandlungen und den verursachten Prüfungskosten zu verzeichnen.

Zusammenfassend ist zu konstatieren, dass sich der Erhalt einer Mandantenerklärung *effektivitätserhöhend* auf die Durchführung von analytischen Prüfungshandlungen auswirkt. Dies gilt

- (1) für Nicht-Spezialisten und
- (2) für Prüfungssituationen mit erhöhtem Prüfungsrisiko,

wenn die Nicht-Fehlererklärung des Mandanten der tatsächliche Grund für die festgestellte Abweichung ist.

Demgegenüber übt der Erhalt einer Mandantenerklärung einen *negativen Einfluss* auf

- (1) die *Effizienz* von analytischen Prüfungshandlungen aus, wenn sie von Spezialisten durchgeführt werden und die Nicht-Fehlererklärung des Mandanten der tatsächliche Grund für die festgestellte Abweichung ist,
- (2) die *Prüfungskosten* aus, wenn ein erhöhtes Prüfungsrisiko zu berücksichtigen ist und die Nicht-Fehlererklärung des Mandanten der tatsächliche Grund für die festgestellte Abweichung ist.

Neben der Darstellung der Untersuchungsergebnisse ist weiterhin ein Vergleich der Ergebnisse mit den Erkenntnissen der bisherigen Prüfungsforschung von besonderem Interesse. Insgesamt ist zu konstatieren, dass der überwiegende Teil der aufgestellten Hypothesen nicht bestätigt werden konnte. Damit weicht ein Großteil der Ergebnisse dieser Studie von den Erkenntnissen der bisherigen Prüfungsforschung ab. So konnte etwa im Rahmen der Hypothesengenerierung überwiegend kein spreading effect und durchweg kein interference effect nachgewiesen werden. In der empirischen Prüfungsforschung konnte ein spreading effect indes regelmäßig und ein interference effect zum Teil fest-

gestellt werden.<sup>1</sup> Bezüglich der Informationssuche ließ die vorliegende Studie überwiegend keine positive Suchstrategie und keine konfirmatorischen Prozesse erkennen. Diese Verzerrungen wurden im Rahmen diverser Studien jedoch mehrfach nachgewiesen.<sup>2</sup> Hinsichtlich der Auswahlhäufigkeit einer Mandantenerklärung konnten innerhalb der vorliegenden Studie ausschließlich bei einer korrekten Mandantenerklärung signifikante Unterschiede nachgewiesen werden, nicht hingegen bei einer inkorrekten Mandantenerklärung. Da in der empirischen Prüfungsforschung bisher keine Studie mit einer korrekten Nicht-Fehlererklärung des Mandanten durchgeführt wurde, kann diesbezüglich keine Gegenüberstellung erfolgen. Allerdings konnte die bisherige Prüfungsforschung bei einer inkorrekten Mandantenerklärung mehrfach eine signifikant höhere Auswahlhäufigkeit feststellen.<sup>3</sup> Insgesamt ist demnach zu konstatieren, dass die Ergebnisse der bisherigen Forschung tendenziell auf einen negativen Einfluss durch den Erhalt einer Mandantenerklärung hinweisen und im Rahmen der vorliegenden Studie indes nur geringfügige negative Auswirkungen und zum Teil sogar positive Auswirkungen zu verzeichnen sind.

Eine Begründung für die unterschiedlichen Ergebnisse könnte etwa in den abweichenden Aufgabenstellungen der Studien liegen. Während die bisherige Prüfungsforschung bis auf wenige Ausnahmen lediglich einzelne Phasen der prüferischen Vorgehensweise betrachtete und die teilnehmenden Probanden somit keine vollständige (analytische) Prüfungshandlung durchführten, wurde in der vorliegenden Studie die Durchführung einer vollständigen (analytischen) Prüfungshandlung verlangt. Dieser höhere Praxisbezug könnte zu einer stärkeren Identifikation und somit Motivation der Teilnehmer geführt haben.

Ein weiterer potenzieller Grund dieser Unterschiede könnte darin liegen, dass die Studien zu verschiedenen Zeitpunkten bzw. -räumen durchgeführt wurden. Während der überwiegende Teil der bisherigen Prüfungsforschung von 1980 bis zur Jahrtausendwende durchgeführt wurde, fand die vorliegende Studie im Laufe des Jahres 2007 statt. Möglicherweise haben modernere Prüfungstechniken bzw. neue Erkenntnisse bzgl. einer optimierten prüferischen Vorgehensweise die Qualität der Jahresabschlussprüfung erhöht.

---

<sup>1</sup> Vgl. Kap. 4.2.1.

<sup>2</sup> Vgl. Kap. 4.2.2.1 sowie Kap. 4.2.2.2.

<sup>3</sup> Vgl. Kap. 0.



Ein nicht zu vernachlässigender Aspekt ist in diesem Zusammenhang auch darin zu sehen, dass der überwiegende Teil der bekanntgewordenen Fälle von Bilanzmanipulationen aus den Jahren nach 2000 stammen. Dieses könnte dazu geführt haben, dass die professionelle Skepsis der Prüfer gegenüber festgestellten Abweichungen im Jahresabschluss und insbesondere gegenüber Nicht-Fehlererklärungen des Mandanten geschärft wurde. Dieses würde wohl eine geminderte Beeinflussbarkeit der prüferischen Vorgehensweise mit sich bringen. Zudem hat sich der Berufsstand in den letzten zehn Jahren mehrfach reformiert. Hinzu kamen diverse Gesetzesinitiativen, die die Aufsicht über und Kontrolle von Berufsständige(n) verschärft haben. Diese Maßnahmen könnten ebenfalls dazu geführt haben, dass sich die Qualität der Abschlussprüfung erhöht hat. Weniger wahrscheinlich erscheint dagegen, dass die Zurkenntnisnahme der Ergebnisse empirischer Studien zu einer erhöhten Wachsamkeit der Prüfer gegenüber Beeinflussungsmöglichkeiten geführt hat. Dieses ist darauf zurückzuführen, dass die überwiegend aus dem angloamerikanischen Sprachraum hervorgegangenen Studien in Deutschland kaum publiziert wurden.

Darüber hinaus könnten die Unterschiede auf Divergenzen in den demographischen Faktoren der Probanden zurückgeführt werden. Diese könnten etwa durch unterschiedliche Verfahren zur Auswahl der potenziellen Teilnehmer hervorgerufen worden sein. Die Auswahl unterlag in der Regel einer Kontaktperson aus der jeweiligen Prüfungsgesellschaft und konnte somit nicht vom Leiter des jeweiligen Experiments kontrolliert werden.

In diesem Zusammenhang ist auch darauf hinzuweisen, dass ein Teil der empirischen Studien im Rahmen von internen Schulungsmaßnahmen der jeweiligen Prüfungsgesellschaft durchgeführt wurde. Demgegenüber nahmen die Prüfer in der vorliegenden Studie überwiegend während ihrer Arbeitszeit teil. Möglicherweise besteht ein Zusammenhang zwischen der Qualifikation von Prüfern und der Wahrscheinlichkeit, mit der Prüfer von ihren Vorgesetzten zur Teilnahme an Weiterbildungsmaßnahmen aufgefordert werden.

Schließlich könnte der Grund für die divergierenden Ergebnisse auch in den unterschiedlichen Erhebungsorten der Studien liegen. Wie bereits zuvor erwähnt stammt ein Großteil der bisherigen Prüfungsforschung aus den angloamerikanischen Ländern. Demgegenüber wurde die vorliegende Studie in Deutschland durchgeführt. Da der Zugang zum Beruf des Wirtschaftsprüfers in Deutschland an wesentlich höheren fachlichen und persönlichen Voraussetzungen gebunden ist als in den angloamerikanischen

Ländern, könnten die unterschiedlichen Ergebnisse auch auf Differenzen in der Qualifikation der Probanden zurückgeführt werden.

#### **4.7 Grenzen der Untersuchung**

Die vorgetragenen Ergebnisse sind vor dem Hintergrund diverser Begrenzungen zu betrachten, die im Folgenden aufgeführt werden:

- Da es sich bei der Auswahl der Probanden um keine Zufallsauswahl handelt, kann kein Anspruch auf Repräsentativität der Ergebnisse erhoben werden. Die Repräsentativität wird insbesondere dadurch eingeschränkt, dass es sich ausschließlich um Prüfer der Big-Four-Prüfungsgesellschaften handelte. Zudem besteht die Möglichkeit, dass bei der Verteilung der Fallstudien durch die interne Kontaktperson überwiegend zuverlässige Mitarbeiter ausgewählt wurden, um eine möglichst positive Außendarstellung zu erreichen. Damit unterlägen die Ergebnisse einer zusätzlichen Verzerrung.
- Die durch den Fehler bzw. Nicht-Fehler verursachte Diskrepanz manifestierte sich ausschließlich in der Bruttogewinnmarge. Möglicherweise reagieren Prüfer auf Abweichungen in anderen Kennzahlen oder Jahresabschlusspositionen auf eine andere Art und Weise, so dass die vorliegenden Ergebnisse nicht auf jegliche analytische Prüfungshandlung übertragen werden können.
- Bei der festgestellten Abweichung handelt es sich um eine Überbewertung in der Bruttogewinnmarge. Es ist denkbar, dass Prüfer ein anderes Verhalten bei einer Unterbewertung der Bruttogewinnmarge offenbaren, weil sie möglicherweise eine andere Risikoeinschätzung vornehmen. Sollten Prüfer unterschiedliche Vorgehensweisen bei Über- und Unterbewertungen an den Tag legen, sind die Ergebnisse nicht generalisierbar.
- Die in dem zu bearbeitenden Fall enthaltene Gewinn- und Verlustrechnung wurde nach dem Umsatzkostenverfahren aufgestellt. In der Praxis werden HGB-Abschlüsse jedoch überwiegend nach dem Gesamtkostenverfahren aufgestellt. Zwar gewinnt das Umsatzkostenverfahren im Zuge der zunehmenden Anwendung der internationalen Rechnungslegungsstandards an Bedeutung. Dennoch könnten Prüfer in der Praxis überwiegend mit dem Gesamtkostenverfahren konfrontiert werden. Folglich wäre die Aussagekraft der vorliegenden Ergebnisse stark eingeschränkt.

- Bei dem implementierten Fehler handelte es sich um eine inkorrekte Verteilung der Standardkostenabweichung. Dieser Fehler tritt in der Prüfungspraxis mit geringer Häufigkeit auf und ist schwierig zu identifizieren.<sup>1</sup> Möglicherweise ist die Erfolgsquote der Prüfer höher, wenn es sich um einen Fehler mit höherer Eintrittshäufigkeit handelt. Insofern können die vorliegenden Ergebnisse nicht auf alle analytischen Prüfungshandlungen übertragen werden.
- In der vorliegenden Studie wurden die teilnehmenden Prüfer ausdrücklich darauf hingewiesen, die Abweichung in der Bruttogewinnmarge zu fokussieren. In der Prüfungspraxis unterliegt es der Eigenverantwortung des Prüfers, einen Gegenstand für den Soll-Ist-Vergleich zu bestimmen. Möglicherweise hat die Vorgabe bereits einen Einfluss auf das Verhalten der Probanden ausgeübt, so dass die Vergleichbarkeit mit der Prüfungspraxis nicht gegeben ist.
- Die Art und der Umfang der möglichen Tests war vorgegeben. Möglicherweise sind nicht alle Tests berücksichtigt worden, die eine Aufdeckung der Abweichungsursache gewährleisten hätten. Insofern ist auch hier eine mangelnde Generalisierbarkeit der Ergebnisse festzuhalten.
- Die Fallstudie wurde von den Probanden extern bearbeitet, so dass keine Beaufsichtigung durch einen Versuchsleiter gewährleistet werden konnte. Es besteht die Möglichkeit, dass die Fallstudien durch Teamarbeit gelöst wurden oder dass Absprachen zwischen den einzelnen Teilnehmern stattfanden. Dadurch wären die Daten erheblich verzerrt.
- Schließlich wurde die Fallstudie im Zeitraum der Busy Season bearbeitet. Möglicherweise hat der hiermit verbundene Zeitmangel der teilnehmenden Prüfer dazu geführt, dass die Bearbeitung mit geringer Sorgfalt durchgeführt wurde. Hierdurch wäre die Repräsentativität der Ergebnisse in Frage gestellt.

---

<sup>1</sup> Vgl. COACKLEY/LOEBBECKE (1985).

## 5 Zusammenfassung und Fazit

Die Besonderheit von analytischen Prüfungshandlungen liegt darin, dass sie die Unternehmensdaten aus einem anderen Blickwinkel betrachten als Einzelfallprüfungen. Letztere zeichnen sich insbesondere dadurch aus, dass sie einzelne Geschäftsvorfälle und Bestände fokussieren und ihre korrekte Abbildung in der Finanzbuchführung des Unternehmens überprüfen. Analytische Prüfungshandlungen hingegen lösen sich von dieser Einzelfallbetrachtung und konzentrieren sich auf das *Gesamtbild* in den Unternehmensdaten, indem sie *funktionale Zusammenhänge* und damit *vorhersehbare Konstellationen und Muster im Jahresabschluss* unterstellen. Weichen die vermuteten Jahresabschlusswerte von den Buchwerten des Unternehmens ab, so könnte der Prüfer dieses zum Anlass nehmen, um die Normenkonformität des Abschlusses in Frage zu stellen. Kann hingegen keine wesentliche Differenz festgestellt werden, könnte dies als ein Indiz für die Abwesenheit wesentlicher Fehler im Jahresabschluss aufgefasst werden.

Im Rahmen der vorliegenden Arbeit wurde zunächst aufgezeigt, dass analytische Prüfungshandlungen immense *Kosteneinsparungspotenziale* besitzen. Dabei kann ihre Effizienz in allen drei Phasen des Prüfungsprozesses zur Geltung kommen. Bei der *Prüfungsplanung* etwa ermöglichen analytische Prüfungshandlungen dem Prüfer besonders risikoorientiert vorzugehen. D.h. personelle und zeitliche Ressourcen werden insbesondere dort eingesetzt, wo analytische Prüfungshandlungen ein erhöhtes Fehlerrisiko indizieren. Prüffelder mit geringerem Fehlerrisiko werden hingegen weniger intensiv geprüft. In der Phase der *Prüfungsdurchführung* können analytische Prüfungshandlungen aussagekräftige Prüfungsnachweise erbringen und dadurch den Einsatz von kosten- und zeitintensiven Einzelfallprüfungen reduzieren. Im Rahmen der *abschließenden Gesamtdurchsicht* stellen analytische Prüfungshandlungen neben dem working paper review die einzige Prüfungsmethode zur Beurteilung der Normenkonformität des Jahresabschlusses dar, da in dieser Prüfungsphase zumeist keine oder lediglich eingeschränkte Möglichkeiten zur Durchführung von weiteren Einzelfallprüfungen bestehen.

Neben der *Effizienz* von analytischen Prüfungshandlungen ist auch ihre *Effektivität* zu betonen. Mitunter sind analytische Prüfungshandlungen die einzige Möglichkeit des Prüfers, einen Hinweis auf die Existenz wesentlicher Fehler im Jahresabschluss zu erhalten. Beispiele für Sachverhalte, bei denen Einzelfallprüfungen versagen, sind etwa die Prüfung der Vollständigkeit von Verbindlichkeiten oder die Aufdeckung von fiktionalen Umsätzen, welche durch gefälschte Ausgangsrechnungen ausreichend dokumentiert werden.

Trotz ihrer Fähigkeit, einen Beitrag zur Steigerung der Qualität von Jahresabschlussprüfungen zu leisten, konnte gezeigt werden, dass analytische Prüfungshandlungen unterschiedlichen Schwächen unterliegen. Ein Nachteil liegt etwa darin, dass ihre Effektivität und Effizienz von den individuellen Fähigkeiten des Abschlussprüfers (sog. *interne Einflussfaktoren*) abhängen. Prüfer müssen nicht nur die unterschiedlichen Verfahren analytischer Prüfungshandlungen kennen. Vielmehr ist es notwendig, dass sich Prüfer auch darüber bewusst sind, *wann* und *wo* ihr Einsatz einen höchstmöglichen Nutzen stiftet. Neben der in abhängig zur Qualität des internen Kontrollsystems stehenden Zuverlässigkeit der eingehenden Daten hat der Prüfer den Aggregationsgrad, die Verlässlichkeit der Datenquelle, die Relevanz und die Vergleichbarkeit der Daten zu berücksichtigen. Darüber hinaus sind analytische Prüfungshandlungen nicht für jegliche Prüffelder in gleicher Weise geeignet. So weisen Positionen der Gewinn- und Verlustrechnung häufiger eine höhere Eignung für analytische Prüfungshandlungen auf als die z.T. durch erhebliche Ermessensspielräume gekennzeichneten Bilanzpositionen. Schließlich setzt der erfolgreiche Einsatz von analytischen Prüfungshandlungen weiterhin voraus, dass der Prüfer über ein nicht unerhebliches Maß an rechnungslegungsspezifischem Wissen verfügt. Da funktionale Zusammenhänge im Datenmaterial z.T. erheblich vom jeweilig zu prüfenden Unternehmen variieren können, sind darüber hinaus mandanten- sowie branchenspezifische Kenntnisse unabdingbar. Da die Beurteilung von wesentlichen unerwarteten Abweichungen in den Jahresabschlussdaten letztendlich auch von dem Wissen des Prüfers über mögliche Ursachen einer Abweichung abhängig ist, setzt der erfolgreiche Einsatz von analytischen Prüfungshandlungen schließlich auch ein hinreichendes Maß an fehlerspezifischem Wissen voraus.

Neben diesen subjektiven von der Person des Prüfers abhängigen Einflussfaktoren ist die Effektivität und Effizienz von analytischen Prüfungshandlungen zudem von prüfungsprozessimmanenten Faktoren (sog. *externe Einflussfaktoren*) abhängig. Die prüferische Urteilsbildung kann etwa durch die im Rahmen eines Prüfungshinweises (z.B. im Prüfungshandbuch der Prüfungsgesellschaft oder im heranzuziehenden Prüfungsstandard) verwendete *Wortwahl*, durch die Kenntnis des *ungeprüften Buchwerts* oder den *Erhalt einer Abweichungserklärung* durch eine dritte Instanz (etwa der Mandant bzw. seine Mitarbeiter, ein Vorgesetzter des Prüfers oder eine EDV-gestützte Entscheidungshilfe) wesentlich beeinflusst werden. Dabei kann Art und Umfang der Beeinflussung etwa durch die Verständlichkeit der „vererbten“ Abweichungsursache oder durch die Anzahl der erhaltenen Abweichungserklärungen variieren. Ferner können sich Gruppenprozesse (etwa die im Rahmen eines Review-Prozesses anfallende Kommunikatio-

nen mit dem Reviewer oder der Interaktion mit anderen Mitgliedern des Prüfungsteam) nachhaltig auf die prüferische Vorgehensweise auswirken. Schließlich kann das Ergebnis der analytischen Prüfungshandlung auch von dem Rechtfertigungs- und Zeitdruck sowie von der Pflicht zur Dokumentation oder Quantifizierung einer Abweichungshypothese nachhaltig beeinflusst werden.

Diese Abhängigkeit analytischer Prüfungshandlungen von internen und externen Faktoren zeigt, dass ihre Vorteile erheblichen Nachteilen gegenüberstehen. Werden analytische Prüfungshandlungen nicht korrekt angewandt oder beeinträchtigen äußere Umstände die Güte der analytischen Prüfungshandlungen, so könnte sich dies negativ auf die Qualität des gesamten Prüfungsprozesses auswirken. Allerdings konnte festgestellt werden, dass das hohe Anspruchsniveau zur Durchführung von analytischen Prüfungshandlungen nicht im Einklang mit der geringen Aufmerksamkeit steht, die ihnen sowohl in der Literatur als auch in den entsprechenden Prüfungsstandards entgegengebracht wird. Aufgrund der mangelnden Strukturierung lag das Ziel der vorliegenden Arbeit darin, eine möglichst hohe Anzahl von Handlungsempfehlungen zur Durchführung von analytischen Prüfungshandlungen herauszuarbeiten. Im Mittelpunkt stand neben einer ausführlichen Erläuterungen der unterschiedlichen Verfahren insbesondere die Untersuchung von potenziellen Faktoren, die sich nachhaltig auf die prüferische Urteilsbildung bei der Durchführung von analytischen Prüfungshandlungen auswirken. Dabei wurde auf die umfassenden Ergebnisse der überwiegend aus dem angloamerikanischen Sprachraum stammenden empirischen Prüfungsforschung zurückgegriffen.

Insgesamt ging aus der Auswertung dieser Forschungsaktivitäten zahlreiche wertvolle Erkenntnisse zur Durchführung von analytischen Prüfungshandlungen hervor. So konnten diverse sog. *kognitive Verzerrungen* identifiziert werden, welche dadurch gekennzeichnet sind, dass sie nicht durch rationales Verhalten erklärt werden und dass sie die Qualität von analytischen Prüfungshandlungen nachhaltig beeinflussen können. Diese Erkenntnisse wurden herangezogen, um auf potenzielle Missstände in der Prüfungspraxis hinzuweisen und gleichzeitig *Handlungsempfehlungen* für eine Vorgehensweise auszusprechen, die den Prüfer weitestgehend vor einer negativen Beeinflussung schützen könnte.

Darüber hinaus wurden Mechanismen untersucht, die kognitiven Verzerrungen entgegenwirken können (sog. *debiasing-Mechanismen*). So konnte etwa gezeigt werden, dass Prüfer in Abhängigkeit der Art und des Umfangs der vorhandenen Erfahrung unterschiedlich stark von äußeren Faktoren beeinflusst wurden. Dieses wurde zum Anlass

genommen, um Handlungsempfehlungen hinsichtlich der Personalzuordnung, der Personalaus- und -weiterbildung auszusprechen. Es konnte ferner identifiziert werden, dass zahlreichen kognitiven Verzerrungen bereits dadurch entgegengewirkt werden konnten, dass den Prüfern unterschiedliche Aufforderungen oder Hinweise im Rahmen der Durchführung von analytischen Prüfungshandlungen mitgeteilt wurden. Da bei einem Großteil der Prüfungsgesellschaften regelmäßig prozessbegleitende Softwarepakete zum Einsatz kommen, wurden zahlreiche Handlungsempfehlungen zur Implementierung derartiger Aufforderungen und Hinweise ausgesprochen.

Eine dominierende Rolle in der o.a. empirischen Prüfungsforschung nimmt der potenzielle Einfluss des Erhalts einer Mandantenerklärung ein, falls die analytische Prüfungshandlung eine wesentliche unerwartete Abweichung anzeigt. Die Brisanz der Frage nach der Zweckmäßigkeit einer derartigen Vorgehensweise resultiert in erster Linie daraus, dass IDW PS 312.27 und ISA 520.18 die Einholung einer Mandantenerklärung empfehlen. Diese Frage konnte trotz des enormen Forschungsaufwands jedoch nicht abschließend geklärt werden, was insbesondere durch die z.T. widersprüchlichen Forschungsergebnisse dokumentiert wird. Daher wurde eine eigene empirische Studie in Form eines Experiments konzipiert und durchgeführt, deren Ziel darin lag, die Ergebnisse der bisherigen Forschungsaktivitäten zu erweitern und zu vertiefen.

Im Ergebnis konnte der Studie entnommen werden, dass die Zweckmäßigkeit der Einholung einer Mandantenerklärung differenziert zu betrachten ist. Insbesondere konnte festgestellt, dass lediglich Prüfer mit geringer branchenspezifischer Erfahrung von dem Erhalt einer Mandantenerklärung profitierten, da sie hierdurch eine höhere Effektivität erreichten. Bei Prüfern mit hoher branchenspezifischer Erfahrung übte der Erhalt einer Mandantenerklärung hingegen keinen Einfluss auf die Effektivität der analytischen Prüfungshandlung aus. Dem steht gegenüber, dass Prüfer mit hoher branchenspezifischer Erfahrung bei gleicher Effektivität erhebliche Effizienzeinbußen erleiden, wenn ihnen eine Mandantenerklärung zur Verfügung gestellt wird.

Vergleichbare Ergebnisse ergaben sich, wenn nicht nach der branchenspezifischen Erfahrung der Prüfer, sondern nach dem zugrunde liegenden Prüfungsrisiko differenziert wurde. Bei erhöhtem Prüfungsrisiko erreichten die Prüfer mit Mandantenerklärung zwar eine höhere Effektivität, allerdings erlangten sie diese nur unter einer Erhöhung der Prüfungskosten, da der Prüfungsumfang wesentlich erhöht wurde.

In Bezug auf die Frage nach der Adäquanz der Einholung einer Mandantenerklärung ist insgesamt ein trade off zwischen der Effektivität und der Effizienz von analytischen

Prüfungshandlungen zu verzeichnen, bei dem der Erhalt einer Mandantenerklärung zwar zu einer höheren Effektivität und somit besseren Fehleraufdeckungsquote führt, gleichzeitig jedoch eine geringere Effizienz erreicht wird. Es ist also abzuwägen, welchem Oberziel der Jahresabschlussprüfung (Wirksamkeit oder Wirtschaftlichkeit der Jahresabschlussprüfung) eine höhere Bedeutung beigemessen wird. Berücksichtigt man jedoch, dass die Jahresabschlussprüfung aufgrund der in den letzten Jahren vermehrt bekanntgewordenen Fälle von Bilanzmanipulation einen erheblichen Imageverlust erlitt und damit das Vertrauen der Öffentlichkeit in die Tätigkeit des Abschlussprüfers gefährdet ist, so sollte nach Auffassung des Autors in erster Linie eine möglichst hohe Wirksamkeit von analytischen Prüfungshandlungen und damit der gesamten Jahresabschlussprüfung gewährleistet werden. Daher spricht sich der Autor für ein Beibehalten der in IDW PS 312.27 und ISA 520.18 empfohlenen Einholung einer Mandantenerklärung aus.



## Literaturverzeichnis

- ALBERT, HANS (1964): Problem der Theoriebildung, Entwicklung, Struktur und Anwendung sozialwissenschaftlicher Theorien. In: Albert, H. (Hrsg.): Theorie und Realität, Ausgewählte Aufsätze zur Wissenschaftslehre der Sozialwissenschaften, Tübingen, S. 3-70.
- ALBERT, HANS (1972): Konstruktion und Kritik, Hamburg.
- ALBERT, HANS (1977): Individuelles Handeln und soziale Steuerung. Die ökonomische Tradition und ihr Erkenntnisprogramm. In: Lenk, H. (Hrsg.): Handlungstheorien interdisziplinär IV, München, S. 177-225.
- ALBERT, MAX (1996): „Unrealistische Annahmen“ und empirische Prüfung. In: Zeitschrift für Wirtschaft- und Sozialwissenschaften, S. 451-486.
- ALDERMAN, C. WAYNE (1987): An Overview of the Use of Analytical Review Procedures in an Audit. In: The CPA Journal, January, S. 91-93.
- ALLEN, ROBERT D./BEASLEY, MARK S. (1999): Improving Analytical Procedures: A Case of Using Disaggregate Multilocation Data. In: Auditing: A Journal of Practice & Theory, Fall, S. 128-142.
- AMEEN, ELSIE C./STRAWSER, JERRY R. (1994): Investigating the Use of Analytical Procedures: An Update and Extension. In: Auditing: A Journal of Practice & Theory, Fall, S. 69-76.
- ANDERSON, CRAIG A./SECHLER, ELISABETH (1986): Effects of Explanation and counter-explanation on the development and Use of Social theories. In: Journal of Personality and Social Psychology, S. 24-34.
- ANDERSON, CRAIG/LEPPER, MARK/ROSS, LEE (1980): Perseverance of social Theories: the role of Explanations in the Persistence of discredited informations. In: Journal of Personality and Social Psychology, S. 1037-1045.
- ANDERSON, JOHN C./JENNINGS, MARIANNE M./KAPLAN, STEVEN E./RECKERS, PHILIP M.J. (1995): The Effect of Using Diagnostic Decision Aids for Analytical Procedures on Judges' Liability Judgments. In: Journal of Accounting and Public Policy, S. 33-62.
- ANDERSON, JOHN C./KAPLAN, STEVEN E./RECKERS, PHILIP M.J. (1992): The Effects of Output Interference on Analytical Procedures Judgements. In: Auditing: A Journal of Practice & Theory, Fall, S. 1-13.

- ANDERSON, JOHN C./KAPLAN, STEVEN E./RECKERS, PHILIP, M.J. (1997): the effect of interference and availability from hypotheses generated by a decision aid upon analytical procedures judgements. In: Behavioral Research in Accounting, S. 1-20.
- ANDERSON, JOHN C./MORENO, KIMBELY K./MUELLER, JENNIFER M. (2003): The effect of client vs. decision aid as a source of explanations upon auditors' sufficiency judgments: a research note. In: Behavioral Research in Accounting, S. 1-11.
- ANDERSON, JOHN R. (2001): Kognitive Psychologie, 3. Aufl., Heidelberg.
- ANDERSON, JOHN.R. (1987): Skill Acquisition: Compilation of Weak-Method Problem Solutions. In: Psychological Review, S. 192-210.
- ANDERSON, URTON/KADOUS, KATHRYN/KOONCE, LISA (2004): The Role of Incentives to Manage Earnings and Quantification in Auditors' Evaluations of Management-Provided Information. In: Auditing: A Journal of Practice & Theory, March, S. 11-27.
- ANDERSON, URTON/KOONCE, LISA (1995): Explanation as a Method for Evaluating Client-Suggested Causes in Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Fall, S. 124-132.
- ANDERSON, URTON/KOONCE, LISA (1998): Evaluating the Sufficiency of Causes in Auditing Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Spring, S. 1-12.
- ANDERSON, URTON/KOONCE, LISA/MARCHANT, GERRY (1994): The Effect of Source-Competence Information and its Timing on Auditors' Performance of Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Spring, S. 137-148.
- ARBINGER, ROLAND (1997): Psychologie des Problemlösens, Darmstadt.
- ARENS, ALVINA./BEASLEY, MARK S./ELDER, RANDAL J. (2005): Auditing and Assurance Services: An Integrated Approach, 10. Aufl., Englewood Cliffs.
- ARRINGTON, EDWARD C. /HILLISON, WILLIAM/ICERMAN, RODA C. (1983): Research in Analytical Review: The State of the Art. In: Journal of Accounting Literature, S. 153-182.
- ASARE, STEPHAN K./TROMPETER, GREGORY M./WRIGHT, ARNOLD M. (2000): The Effect of Accountability and Time Budgets on Auditors' Testing Strategies. In: Contemporary Accounting Research, S. 539-560.

- ASARE, STEPHAN K./WRIGHT, ARNOLD M. (1997): Hypothesis Revision Strategies in Conducting Analytical Procedures. In: Accounting, Organizations and Society, S. 737-755.
- ASARE, STEPHAN K./WRIGHT, ARNOLD M. (2001): Design Considerations for Research on Analytical Procedures. In: International Journal of Auditing, S. 205-214.
- ASARE, STEPHEN K./WRIGHT, ARNOLD M. (1997): Evaluation of Competing Hypotheses in Auditing. In: Auditing: A Journal of Practice & Theory, Spring, S. 1-11.
- ASARE, STEPHEN K./WRIGHT, ARNOLD M. (2003): A Note on the Interdependence between Hypothesis Generation and Information Search in Conducting analytical Procedures. In: Contemporary Accounting Research, S. 235-251.
- ASARE, STEPHEN K./WRIGHT, ARNOLD/WRIGHT, SALLY (1998): Utilizing Analytical Procedures as Substantive Evidence: The Impact of a Client Explanation on Hypothesis Testing. In: Hunton, J. (Hrsg.): Advances in Accounting Behavioral Research, Stamford. S. 13-32.
- ASHTON, ALISON H. (1991): Experience and error frequency knowledge as potential determinants of audit expertise. In: The Accounting Review, S. 218-239.
- ASHTON, ALISON H./ASHTON, ROBERT H. (1988): Sequential Belief Revision in Auditing. In: The Accounting Review, S. 623-641.
- ASHTON, ROBERT H. (1982): Human Information Processing in Accounting, Sarasota.
- ASHTON, ROBERT H./ASHTON, ALISON H. (1995): Perspectives on Judgment and Decision-Making Research in Accounting and Auditing. In: Ashton, R./Ashton, A. (Hrsg.): Judgment and Decision-Making Research in Accounting and Auditing., New York, S. 3-25.
- ATTESLANDER, PETER (1993): Methoden der empirischen Sozialforschung, 7. Aufl., Berlin.
- ATTESLANDER, PETER (2006): Methoden der empirischen Sozialforschung, 11. Aufl., Berlin.
- ATTESLANDER, PETER/KOPP, MANFRED (1995): Befragung. In: Roth, E. (Hrsg.): Sozialwissenschaftliche Methoden, München , Wien, S. 146-174.

- AWASHTI, VIDYA/PRATT, JAMIE (1990): The Effect of Monetary Incentives on Effort and Decision Performance: The Role of Cognitive Characteristics. In: The Accounting Review, S. 797-811.
- AYERS, SUSAN/KAPLAN, STEVEN E. (1993): An Examination of the Effect of Hypothesis Framing on Auditors' Information Choices in an Analytical Procedure Task. In: ABACUS, S. 113-130.
- BACKHAUS, KLAUS (2003): Industriegütermarketing, 7. Aufl., München.
- BACKHAUS, KLAUS/ERICHSON, BERND/PLINKE, WULFF/WEIBER, ROLF (2006): Multivariate Analysemethoden – Eine Anwendungsorientierte Einführung, 11. Aufl., Berlin.
- BAETGE, JÖRG/KIRSCH, HANS-JÜRGEN/THIELE, STEFAN (2004): Bilanzanalyse, 2. Aufl., Düsseldorf.
- BAETGE, JÖRG/KIRSCH, HANS-JÜRGEN/THIELE, STEFAN (2005): Bilanzen, 8. Aufl., Düsseldorf.
- BAILEY, ANDREW D. (1981): Statistical Auditing, New York.
- BALLOU, BRIAN (2001): The Relationship between Auditor Characteristics and the Nature of Review Notes for Analytical Procedure Working Papers. In: Behavioural Research in Accounting, S. 25-48.
- BAMBER, MICHAEL E/GILLET, PETER R./MOCK, TED J./TROTMAN, KEN T. (1995): Audit Judgment. In: Bell, T.B./Wright, A.M. (Hrsg.): Auditing Practice, Research and Education: A Productive Collaboration, New York, S. 55-85.
- BAMBER, GÜNTER/BAUR, FRANZ (2002): Statistik, 12. Aufl., München.
- BECKER, WILHELM/PETERSON, HEINZ (1964): Entwicklungstendenzen auf dem Gebiet der Prüfungsmethoden. In: Die Wirtschaftsprüfung, S. 408-415.
- BEDARD, JEAN C./BIGGS, STANLEY F. (1991): Pattern Recognition, Hypotheses Generation, and Auditor Performance in an Analytical Task. In: The Accounting Review, S. 622-642.
- BEDARD, JEAN C./BIGGS, STANLEY F. (1991a): Pattern Recognition, Hypotheses Generation, and Auditor Performance in an Analytical Task. In: The Accounting Review, S. 622-642.
- BEDARD, JEAN C./BIGGS, STANLEY F. (1991b): The Effect of Domain-Specific Experience on Evaluation of Management Representations in Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Supplement., S. 78-90.

- BEDARD, JEAN C./BIGGS, STANLEY F./MARONEY, JAMES J. (1998): Sources of Process Gain and Loss From Group Interaction in Performance of Analytical Procedures. In: Behavioural Research in Accounting, S. 207-233.
- BEDARD, JEAN C./WRIGHT, ARNOLD M. (1994): The Functionality of Decision Heuristic: Reliance on Prior Audit Adjustments in Evidential Planning. In: Behavioral Research in Accounting, S. 64 - 89.
- BEDARD, JOHN C. (1989): An archival investigation of audit program planning. In: Auditing: A Journal of Practice & Theory, Spring, S. 57-71.
- BEHRENS, GEROLD (1993): Wissenschaftstheorie und Betriebswirtschaftslehre. In: Wittmann, W. (Hrsg.): HWB der Betriebswirtschaft; 5. Aufl., Sp. 4763-4772.
- BERNSTEIN, LEOPOLD A. (1986): Analytical Auditing. In: Cashin, J.A./Neuwirth, P.D./Levy, J.F. (Hrsg.): Cashin's Handbook for Auditors, 2. Aufl., New York, S. 41/1-41/20.
- BHATTACHARJEE, SUDIP/KIDA, THOMAS/HANNO, DENNIS M. (1999): The Impact of Hypothesis Set Size on the Time Efficiency and Accuracy of Analytical Review Judgments. In: Journal of Accounting Research, S. 83-100.
- BHATTACHARYA, SOMNATH (1994): The Use of Recall Recognition Techniques in Eliciting Experienced-Inexperienced Auditor Differences in the Detection of Management Fraud, Dissertation, University of South Florida.
- BIDDLE, GARY C./JOYCE, EDWARD J. (1982): Heuristics and Biases: Some Implications for Probabilistic Inference in Auditing. In: Symposium on Auditing Research IV, University of Illinois at Urbana-Champaign, 1982, S. 164-194.
- BIERSTAKER, JAMES L./BEDARD, JEAN C./BIGGS, STANLEY F. (1999): The Role of Problem Representation Shifts in Auditor Decision Processes in Analytical Procedures. In: Auditing, A Journal of Theory & Practice, Spring, S. 18-36.
- BIERSTAKER, JAMES L./WRIGHT, SALLY (2001): A Research Note concerning Practical Problem-Solving Ability as a Predictor of Performance in Auditing Tasks. In: Behavioral Research in Accounting, S. 49 - 62.
- BIGGS, STANLEY F./MOCK, THEODORE J./WATKINS, PAUL R. (1988): Auditor's Use of Analytical Review in Audit Program Design. In: The Accounting Review, S. 148-161.

- BIGGS, STANLEY F./WILD, JOHN J. (1984): A Note on the Practice of Analytical Review. In: Auditing: A Journal of Practice & Theory, Spring, S. 68-79.
- BIGGS, STANLEY F./WILD, JOHN J. (1985): An Investigation of Auditors Judgement in Analytical Review. In: The Accounting Review, S. 607-633.
- BIGGS, STANLEY F/MOCK, THEODORE J./QUICK, REINER (2000): Das Prüfungsurteil bei analytischen Prüfungshandlungen – Praktische Implikationen von Forschungsergebnissen. In: Die Wirtschaftsprüfung, S. 169-178.
- BIGGS, STANLEY. F./MOCK, THEODORE J. (1983): An Investigation of Auditor Decision Processes in The Evaluation of Internal Controls and Audit Scope Decisions. In: Journal of Accounting Research, S. 234-255.
- BLOCHER, EDWARD (1983): Approaching Analytical Review. In: The CPA Journal, March, S. 24-32.
- BLOCHER, EDWARD/COOPER, JEAN C. (1988): A Study of Auditors' Analytical Review Performance. In: Auditing: A Journal of Practice & Theory, Spring, S. 1-28.
- BLOCHER, EDWARD/KRULL, GEORGE W. JR./TASHMAN, LEONARD J./YATES, STEPHAN V. (2002): Updating Analytical Procedures. In: The CPA Journal, November, S. 37-41.
- BLOCHER, EDWARD/PATTERSON, GEORGE F. JR. (1996): The Use of Analytical Procedures. In: Journal of Accountancy, February, S. 53-55.
- BLOCHER, EDWARD/WILLINGHAM, JOHN J. (1984): Analytical Review: A Guide to Evaluating Financial Statements, New York.
- BLOCHER, EDWARD/WILLINGHAM, JOHN J. (1985): Analytical Review, New York.
- BODE, JÜRGEN (1997): Der Informationsbegriff in der Betriebswirtschaftslehre. In: Zeitschrift für betriebswirtschaftliche Forschung, S. 449-468.
- BONNER, SARAH E./PENNINGTON, NANCY (1991): Cognitive Processes and Knowledge as Determinants of Auditor Expertise. In: Journal of Accounting Literature, S. 1-50.
- BONNER, SARAH E./WALKER PAUL L. (1994): The effect of instructions and experience on the acquisition of auditing knowledge. In: The Accounting Review, S. 157-178.
- BORTZ, JÜRGEN (2005): Statistik für Human- und Sozialwissenschaftler, Heidelberg.
- BORTZ, JÜRGEN/DÖRING, NICOLA (2002): Forschungsmethoden und Evaluation, Berlin.
- BOX, GEORGE E.P./JENKINS, GWILYM M. (1976): Time Series Analysis – Forecasting and Control, San Francisco.

- BRANDER, SYLVIA/KOMPA, AIN/PELTZER, ULF (1989): Denken und Problemlösen – Einführung in die kognitive Psychologie, 2. Aufl., Opladen.
- BROMME REINER/HÖMBERG ECKHARD (1977): Psychologie und Heuristik, Darmstadt.
- BRONNER, ROLF/WITTE, EBERHARD/WOSSIDLO, PETER R. (1972): Betriebswirtschaftliche Experimente zum Informationsverhalten in Entscheidungsprozessen. In: Witte, E. (Hrsg.): Das Informationsverhalten in Entscheidungsprozessen, Tübingen, S. 165-217.
- BROWN, CLIFTON E./PEECHER, MARK E./SOLOMON, IRA (1999): Auditors' Hypothesis Testing in Diagnostic Inference Tasks. In: Journal of Accounting Research, S. 1-25.
- BUCHNER, ROBERT (1997): Wirtschaftliches Prüfungswesen, 2. Aufl., München.
- BU-PEOW NG, TERENCE/GREEN, WENDY/SIMNETT, ROGER (2001): The Effects of Fraud Risk and Management Representation on Auditors' Hypothesis Generation. In: ABACUS, S. 352-368.
- BUTT, JANE L. (1988): Frequency Judgment in an Audit-Related Task. In: Journal of Accounting Research, S. 315-330.
- CALDERON, THOMAS G./GREEN, BRIAN P. (1994): Signalling fraud by using analytical procedures. In: The Ohio CPA Journal, April, S. 27-38.
- CALLAHAN, PATRICK S./JAENICKE, HENRY R./NEEBES, DONALD L. (1988): SASs nos. 56 and 57: Increasing Audit Effectiveness. In: Journal of Accountancy, October, S. 56-68.
- CAPLAN, DENNIS (1999): Internal Controls and the Detection of Management Fraud. In: Journal of Accounting Research, S. 101-117.
- CASHIN, JAMES A./NEUWIRTH, PAUL D./LEVY, JOHN F. (1986): Cashin's Handbook for Auditors, 2. Aufl., New York.
- CASTER, PAUL/PINCUS, KAREN V. (1996): An Empirical Test of Bentham's Theory of the Persuasiveness of Evidence. In: Auditing: A Journal of Practice & Theory, Spring, S. 1-22.
- CHANG, JANIE C./HO, JOANNA L.Y./LIAO, WOODY M. (1997): The Effect of Justification, Task Complexity and Experience/Training on Problem-Solving Performance. In: Behavioral Research in Accounting, S. 98-116.

- CHANG, JANIE C./YEN, SIN-HUI/DUH, RONG-RUEY (2002): An Empirical Examination of Competing Theories to Explain the Framing Effect in Accounting-Related Decisions. In: *Behavioral Research in Accounting*, S. 35-64.
- CHASE, WILLIAM G./SIMON, HERBERT A. (1973): The Minds Eye in Chess. In: William C. (Hrsg.): *Visual Information Processing*, New York, S. 215-281.
- CHEN, YINING/LEITCH, ROBERT A. (1998): The error detection of structural analytical procedures: A simlation study. In: *Auditing: A Journal of Practice & Theory*, Fall, S. 36-70.
- CHEN, YINING/LEITCH, ROBERT A. (1999): An Analysis of the Relative Power Characteristics of Analytical Procedures. In: *Auditing: A Journal of Practice & Theory*, Fall, S. 35-69.
- CHMIELEWICZ, KLAUS (1994): *Forschungskonzeption der Wirtschaftswissenschaft*, 3. Aufl., Stuttgart.
- CHOO, FREDDIE (1989): Cognitive Scripts in Auditing and Accounting Behavior. In: *Accounting, Organizations and Society*, S. 481-493.
- CHRIST, MARY Y. (1993): Evidence on the Nature of Audit Planning Problem Representations: An Examination of Auditor Free Recalls. In: *The Accounting Review*, S. 304-322.
- CHURCH, BRYAN K. (1990): Auditors' use of Confirmatory Processes. In: *Journal of Accounting Literature*, S. 81-112.
- CHURCH, BRYAN K. (1991): An examination of the effect that commitment to hypothesis has on auditors' evaluations confirming and disconfirming evidence. In: *Contemporary Accounting Research* , S. 513-534.
- CHURCH, BRYAN K./MCMILLAN, JEFFREY J./SCHNEIDER, ARNOLD (2001): Factors Affecting Internal Auditors' Consideration of Fraudulent Financial Reporting during Analytical Procedures. In: *Auditing: A Journal of Practice & Theory*, March, S. 65-80.
- CHURCH, BRYAN K./SCHNEIDER, ARNOLD (1993): Auditor's Generation of Diagnostic Hypotheses in Response to a Superior's Suggestion: Interference Effects. In: *Contemporary Accounting Research*, S. 333-350.
- CHURCH, BRYAN K./SCHNEIDER, ARNOLD (1995): Internal Auditors' Memory for Financial Statement Error. In: *Behavioral Research in Accounting*, S. 17-36.



- CLARKSON, PETER M./EMBY, CRAIG/WATT, VANESSA W-S (2002): Debiasing the Outcome Effect: The Role of Instructions in an Audit Litigation Setting. In: Auditing: A Journal of Practice & Theory, September, S. 7-20.
- COACKLEY, JAMES R./LOEBBECKE, JAMES K. (1985): The Expectation of Accounting Errors in Medium-Sized Manufacturing Firms. In: Advances in Accounting, S. 199-245.
- COENENBERG, ADOLF G. (2005): Jahresabschluss und Jahresabschlussanalyse — Aufgaben und Lösungen, 20. Aufl., Stuttgart.
- COGLITORE, FRANK/BERRYMAN, R. GLEN (1988): Analytical Procedures: A Defensive Necessity. In: Auditing: A Journal of Practice & Theory, Spring, S. 150-163.
- COHEN, JEFFREY R. (1994): Further Evidence of Auditor's Asymmetric Reactions to Analytical Review Results. In: Advances in Accounting, S. 167-185.
- COHEN, JEFFREY R./KRISHNAMOORTHY, GANESH/WRIGHT, ARNOLD M. (2000): Evidence on the Effect of Financial and Nonfinancial Trends on Analytical Review. In: Auditing, Spring, S. 27-48.
- COHEN, JEFFREY/KIDA, THOMAS (1989): The Impact of Analytic Review Results, Internal Control Reliability and Experiences on Auditors' Use of Analytical Review. In: Journal of Accounting Research, S. 263-276.
- CROWDER, NITA (1997): Fraud Detection Techniques. In: Internal Auditor, April, S. 17-20.
- CUSHING, BARRY E. (1977): The Coming of Age of Auditing Research. In: Cushing, B./Krosstad, J. (Hrsg.): Frontiers of Auditing Research, Austin (TX), S. 3-8.
- CUSHING, BARRY E./LOEBBECKE, JAMES K. (1986): Comparison of Audit Methodologies of Large Accounting Firms, Studien in Accounting Research 26, American Accounting Association, Sarasota (FL).
- DAROCA, FRANK P./HOLDER, WILLIAM W. (1985): The Use of Analytical Procedures in Review and Audit Engagement. In: Auditing: A Journal of Practice & Theory, Spring, S. 81-92.
- DAVIS, RANDALL/SHROBE, HOWARD/SZOLOVITS, PETER (1993): What is a Knowledge Representation? In: AI (Artificial Intelligence) Magazine, Spring, S. 17-33.
- DEAKIN, EDWARD B./GRANOF, MICHAEL H. (1974): Regression Analysis as a Means of Determining Audit Sample Size. In: The Accounting Review, S. 764-771.

- DEAKIN, EDWARD B./GRANOF, MICHAEL H. (1976): Directing Audit Effort Using Regression Analysis. In: The CPA Journal, February, S. 29-33.
- DEFLIESE, PHILIP L./JAENICKE, HENRY R./O'REILLY, VINCENT M./HIRSCH, MURRAY B. (1990): Montgomery's Auditing, 11. Aufl., New York et al.
- DÖRNER, DIETER/KREUZIG, HEINZ W./REITHER, FRANK/STÄUDEL, THEA (1994): Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität, Bern et al.
- DÖRNER, DIETRICH (1974): Die kognitive Organisation beim Problemlösen. Versuche zu einer kybernetischen Theorie der elementaren Informationsverarbeitungsprozesse beim Denken, Bern et al.
- DÖRNER, DIETRICH (1987): Problemlösung als Informationsverarbeitung, Stuttgart.
- DREXL, ANDREAS (1990): Planung des Ablaufs von Unternehmensprüfungen, Stuttgart.
- DUGAN, MICHAEL T./GENTRY, JAMES A./SHRIVER, KEITH A. (1985): The X-11 Modell: A New Analytical Review Technique for the Auditor. In: Auditing: A Journal of Practice & Theory, Spring, S. 11-22.
- DUNCKER, KARL (1974): Zur Psychologie des produktiven Denkens, 3. Aufl., Berlin.
- DZENG, SIMON C. (1994): A Comparison of Analytical Procedure Expectation Models Using Both Aggregate and Disaggregate Data. In: Auditing: A Journal of Practice & Theory, Fall, S. 1-24.
- EAGLY, ALICE H./CHAIKEN, SHELLY (1984): Cognitive Theories of Persuasion. In: Advances in Experimental Social Psychology, S. 267-359.
- EGNER, HENNING (1984): Über „grenzüberschreitendes wissenschaftliches Arbeiten“ und die Dilettantismusgefahr. In: Zeitschrift für betriebswirtschaftliche Forschung, S. 421-431.
- EGNER, HENNING (1992): Prüfungstheorie, verhaltensorientierter Ansatz (Syllogistischer Ansatz). In: Coenenberg, A./v.Wysocki, K. (Hrsg.): Handwörterbuch der Revision, 2. Aufl., Stuttgart, S. 1566-1578.
- EINHORN, HILLEL J. (1976): A Synthesis: Accounting and Behavioural Science. In: Journal of Accounting Research, S. 196-206.
- EINHORN, HILLEL J./HOGARTH ROBIN M. (1986): Judging probable cause. In: Psychological Bulletin, S. 3-19.

- ELSCHEN, REINER (1991): Gegenstand und Anwendungsmöglichkeiten der Agency-Theory. In: Zeitschrift für betriebswirtschaftliche Forschung, S. 1002-1012.
- ELSTER, JON (1999): Alchemies of the mind: rationality and the emotions, Cambridge.
- EMBY, CRAIG (1994): Framing and Presentation Mode Effects in Professionals Judgment: Auditors' Internal Control Judgments and Substantive Testing Decisions. In: Auditing: A Journal of Practice & Theory, Supplement, S. 102-115.
- EMBY, CRAIG/FINLEY, DAVID (1997): Debiasing Framing Effects in Auditors' Internal Control Judgments and Testing Decisions. In: Contemporary Accounting Research, S. 55-77.
- EMBY, CRAIG/GELARDI, ALEXANDER M.G./LOWE, D. JORDAN (2002): A Research Note on the Influence of Outcome Knowledge on Audit Partners' Judgments. In: Behavioral Research in Accounting, S. 87-104.
- EPSTEIN, SEYMOR/PACINI, ROSEMARY/DENES-RAJ, VERONIKA/HEIER, HARRIET (1996): Individual differences in intuitive-experiential and analytical-rational thinking styles. In: Journal of Personality and Social Psychology, S. 390-405.
- ERNST, GEORG/NEWELL, ALLEN (1969): GPS: A Case Study in Generality and Problem Solving, New York.
- EWERT, RALF (1993): Rechnungslegung, Wirtschaftsprüfung, rationale Akteure und Märkte: Ein Grundmodell zur Analyse der Qualität von Unternehmenspublikationen. In: Zeitschrift für betriebswirtschaftliche Forschung, S. 712-747.
- EWERT, RALF (1999): Wirtschaftsprüfung und ökonomische Theorie – Ein selektiver Überblick. In: Richter, M. (Hrsg.): Theorie und Praxis der Wirtschaftsprüfung II, Berlin, S. 35-100.
- EWERT, RALF (2002): Der informationsökonomische Ansatz des Controlling. In: Weber, J. (Hrsg.): Controlling als akademische Disziplin, 1. Aufl., Wiesbaden, S. 21-37.
- FEROZ, EHSAN H./KIM, SUNGSOO/RAAB, RAY (2005): Performance measurement in corporate governance: Do mergers improve managerial performance in post-merger period? In: Review of Accounting & Finance, Vol. 4, No. 3, S. 86-100.
- FIEDLER, KLAUS (1980): Urteilsbildung als kognitiver Vorgang, München.
- FINLEY, DAVID R. (1994a): Game Theoretic Analysis of Discovery Sampling for Internal Fraud Control Auditing. In: Contemporary Accounting Research, S. 91-114.

- FINLEY, DAVID R. (1994b): Use of Standards in Auditing for Hidden Actions. In: *Advances in Accounting*, S. 255-279.
- FISCHER WINKELMANN/WOLF F. (1992): Prüfungstheorie, empirisch-kognitiver Ansatz. In: Coenenberg, A./v.Wysocki, K. (Hrsg.): *Handwörterbuch der Revision*, 2. Aufl., Stuttgart, Sp.1532-1544.
- FISCHER-WINKELMANN, WOLF F. (1981): Verhaltenswissenschaftliche Ausrichtung der Wirtschaftsprüfung. In: *Zukunftsaufgaben der Revision*, Zürich, S. 67-87.
- FISCHER-WINKELMANN, WOLF F. (1975): *Entscheidungsorientierte Prüfungslehre*, Berlin.
- FITZSIMONS, ADRIAN P./LEVINE, MARC H. (1989): *WG&L Audit Manual*, Boston, New York.
- FLEMING, JOHN (2004): Analytical Procedures are a Helpful Tool, In: *Pennsylvania CPA Journal*, Winter, S. 3-4.
- FRASER, IAN A. M./HATHERLY, DAVID J./LIN, KENNY Z. (1997): An Empirical Investigation of the Use of Analytical Review by External Auditors. In: *The British Accounting Review*, S. 35-47.
- FREDERICK, DAVID M. (1991): Auditors' Representation and Retrieval of Internal Control Knowledge. In: *The Accounting Review*, S. 240-258.
- FREDERICK, DAVID M./HEIMAN-HOFFMAN, VICKY B./LIBBY, ROBERT (1994): The Structure of Auditors' Knowledge of Financial Statement Errors. In: *Auditing: A Journal of Practice & Theory*, Spring, S. 1-21.
- FULLER, LORI R./KAPLAN, STEVEN E. (2004): A note about the effect of auditor cognitive style on task performance. In: *Behavioral Research in Accounting*, S. 131-143.
- GANS, CHRISTIAN (1986): *Betriebswirtschaftliche Prüfungen als heuristische Suchprozesse*, Bergisch Gladbach et al.
- GÄRTNER, MICHAEL (1994): *Analytische Prüfungshandlungen im Rahmen der Jahresabschlussprüfung: Ein Grundsatz ordnungsmäßiger Abschlussprüfung*, Marburg.
- GAUNT, JAMES E./GLEZEN, G. WILLIAM (1996): Guidance for multi-location audits. In: *The CPA Journal*, January, S. 60-62.
- GAUNTT, JAMES E./GLEZEN, G. WILLIAM (1997): Analytical auditing procedures. In: *Internal Auditor*, February, S. 56-60.

- GAVANSKI, IGOR/HUI, CHUN (1992): Natural sample Spaces and uncertain belief. In: Journal of Personality and Social Psychology, S. 766-780.
- GEMÜNDEN, HANS G. (1993): Informationsverhalten. In: Ergebnisse empirischer betriebswirtschaftlicher Forschung – Zu einer Realtheorie der Unternehmung. In: Hauschildt, J./Grün, O. (Hrsg.): Festschrift für Eberhard Witte, Stuttgart, S. 841-877.
- GILL, GURDARSHAN S./COSSERAT, GRAHAM/LEUNG, PHILOMENA/CORAM, PAUL (2001): Modern Auditing and Assurance Services, 2. Aufl., Brisbane.
- GLOVER, STEVEN M. (1997): The influence of time pressure and accountability on auditors' processing of non-diagnostic information. In: Journal of Accounting Research, S. 213-226.
- GLOVER, STEVEN M./JIAMBALVO, JAMES/KENNEDY, JANE (2000): Analytical Procedures and Audit-Planning Decisions. In: Auditing: A Journal of Practice & Theory, Fall, S. 27-45.
- GLOVER, STEVEN M./PRAWITT, DOUGLAS F./WILKS, T. JEFFREY (2005): Why Do Auditors Over-Rely on Weak Analytical Procedures? The Role of Outcome and Precision. In: Auditing: A Journal of Practice & Theory, Supplement, S. 197-220.
- GRAHM, LYNFORD E. (1981): Analytical Review Techniques: Some Neglected Tools. In: The CPA Journal, October, S. 18-24.
- GRAY, IAIN/MANSON, STUART (2000): The audit process: principles, practice and cases, 2. Aufl., London.
- GREEN, WENDY J. (2004a): Impact of the timing of receipt of an inherited explanation on auditors' analytical procedures judgements. In: Accounting and Finance, S. 369-392.
- GREEN, WENDY J. (2004b): Industry Specialists: Are they more Efficient and Effective in Performing Analytical Procedures than Experienced Non-Specialists?, Working Paper.
- GREEN, WENDY J./TROTSMANN, KENT T. (2003): An Examination of Different Performance Outcomes in an Analytical Procedures Task. In: Auditing: A Journal of Practice & Theory, September, S. 219-235.

- GREENLEE, JANET S./BUKOVINSKY, DAVID (1998): Financial ratios for use in the analytical review of charitable organizations. In: The Ohio CPA Journal, January, S. 32-38.
- GREENO, JAMES G. (1978): Natures of problem solving abilities. In: Estes, W.K. (Hrsg.): Handbook of learning and cognitive processes: Human information processing (Vol. 5), Hillsdale et al., S. 239-270.
- GROCHLA, ERWIN (1978): Einführung in die Organisationstheorie, Stuttgart.
- GRUBER, HANS/MANDEL, HEINZ (1996): Expertise und Erfahrung. In: Gruber, H./Ziegler A. (Hrsg.): Expertiseforschung – Theoretische und methodische Grundlagen, Opladen, S. 18-34.
- GUY, DAN M./WINTERS, ALAN J./ALDERMAN, C. WAYNE (1990): Auditing, 2. Aufl., San Diego.
- GWILLIAM, DAVID (1987): A survey of auditing research, Englewood Cliffs.
- HACKMAN, J. RICHARD/MORRIS, CHARLES G. (1975): Group Tasks, Group Interaction Process and Group Performance Effectiveness: A Review and Proposed Integration. In: Berkowitz, L. (Hrsg.): Advances in Experimental Social Psychology, New York, S. 45-99.
- HALLER, AXEL (1992): Empirische Forschung im Prüfungswesen. In: Coenenberg, A./v.Wysocki, K. (Hrsg.): Handwörterbuch der Revision, 2. Aufl., Stuttgart, Sp. 413-426.
- HARPER, ROBERT M. JR./STRAWSER, JERRY R./TANG, KWEI (1990): Establishing Investigation Thresholds for Preliminary Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Fall, S. 115-133.
- HASTIE, REID. C. (1984): Causes and effects of causal attribution. In: Journal of Personality and Social Psychology, S. 44-56.
- HASTIE, REID/PENROD, STEVEN/PENNINGTON, NANCY (1983): Inside the Jury, Cambridge.
- HAYNES, CHRISTINE M. (1993): An Examination into the Effects of Context and Experience on Auditors' Belief Revisions in Cascaded-Interface Tasks, Dissertation, University of Texas, Austin.
- HEIMAN, VICKY B. (1990): Auditors' Assessments of the Likelihood of Error Explanations in Analytical Review. In: The Accounting Review, S. 875-890.

- HEIMAN-HOFFMAN, VICKY B./MOSER, DONALD V./JOSEPH, JERRY A. (1995): The Impact of an Auditor's Initial Hypothesis on Subsequent Performance at Identifying Actual Errors. In: *Contemporary Accounting Research*, S. 763-779.
- HEINTZ, JAMES A./WHITE, GWENDOLEN B. (1989): Auditor Judgement in Analytical Review - Some Further Evidence. In: *Auditing: A Journal of Practice & Theory*, Spring, S. 22-39.
- HEINTZ, JAMES A./WHITE, GWENDOLEN B./BEDARD, JEAN C. (1999): The Effect of Data Reliability on the Influence of Unaudited Values in Audit Analytical Procedures. In: *International Journal of Auditing*, S. 135-146.
- HERMANSON, DANA R. (1997): Commentary on The Effects of Interference and Availability From Hypotheses Generated by a Decision Aid Upon Analytical Procedures Judgements. In: *Behavioural Research in Accounting*, S. 21-26.
- HESSE, FRIEDRICH W./HAHN, C. (1994): Die Rolle der Kontextähnlichkeit beim analogen Problemlösen. In: *Sprache und Kognition*, S. 90-102.
- HILL, G. (1982): Group versus Individual Performance: Are  $n + 1$  heads better than 1? In: *Psychological Bulletin*, S. 517-539.
- HINSZ, VERLIN B. (1990): Cognitive and Consensus Processes in Group Recognition Memory Performance. In: *Journal of Personality and Social Psychology*, S. 705-718.
- HIRST, D. ERIC (1994): Auditors' sensitivity to source reliability. In: *Journal of Accounting Research*, S. 115-126.
- HIRST, D. ERIC/KOONCE, LISA (1996): Audit Analytical Procedures: A Field Investigation. In: *Contemporary Accounting Research*, S. 457-486.
- HITZIG, NEAL B. (2004): The Hidden Risk in Analytical Procedures: What WorldCom Revealed. In: *The CPA Journal*, February, S. 32-35.
- HO, JOANNA L. (1994): The Effect of Compatibility between self assessment and the provided Hypothesis on Auditors' information search strategy. In: *Advances in Accounting*, S. 209-228.
- HO, JOANNA L./MAY, ROBERT G. (1993): Auditors' Causal Probability Judgements in Analytical Procedures for Audit Planning. In: *Behavioural Research in Accounting*, S. 78-100.
- HOGARTH, ROBIN M. (1991): A Perspective on Cognitive Research in Auditing. In: *The Accounting Review*, S. 277-290.

- HOLDER, WILLIAM W. (1983): Analytical Review Procedures in Planning the Audit: An Application Study. In: Auditing: A Journal of Practice & Theory, Spring, S. 100-107.
- HOLDER, WILLIAM W./COLLMER, SHERYL (1980): Analytical Review Procedures: New Relevance. In: The CPA Journal, November, S. 29-35.
- HUBER, OSWALD (1995): Das psychologische Experiment: Eine Einführung, Bern.
- HUSSY, WALTER (1998): Denken und Problemlösen (Thinking and problem solving), Stuttgart.
- HÜTTNER, MANFRED (1986): Prognoseverfahren und ihre Anwendung, Berlin.
- HYLAS, ROBERT E./ASHTON, ROBERT H. (1982): Audit Detection of Financial Statement Errors. In: The Accounting Review, S. 751-765.
- IDW (2006): WP Handbuch 2006, Wirtschaftsprüfung, Rechnungslegung, Beratung, Band 1, 13. Aufl., Düsseldorf.
- ISAAC, DOUG (1996): Discussion of Measuring Cognitive Effort During Analytical Review: A Process-Tracing Framework with Experimental Results. In: Auditing: A Journal of Practice & Theory, Supplement, S. 115-117.
- ISMAIL, ZUBAIDAH/TROTMAN, KEN T. (1995): The Impact of the Review Process in Hypothesis Generation Tasks. In: Accounting, Organizations and Society, S. 345-357.
- ISMAIL, ZUBAIDAH/TROTMAN, KEN T. (1999): Audit Review Effectiveness in an Analytical Review Task. In: International Journal of Auditing, S. 21-28.
- JAMAL, KARIM/JOHNSON, PAUL E./BERRYMAN, GLEN R. (1995): Detecting Framing Effects in Financial Statements. In: Contemporary Accounting Research, S. 85-105.
- JOHNSON, ERIC N. (1994): Auditor Memory for Audit Evidence: Effects of Group Assistance, Time Delay and Memory Task. In: Auditing: A Journal of Practice & Theory, Spring, S. 36-56.
- JOHNSON, LAURENCE E./JOHNSON, ERIC N. (1997): Differences in Planning-Phase Analytical Procedures Between Municipal and Commercial Clients: Initial Evidence. In: Journal of Applied Business Research, Vol. 13, No. 2, S. 37-45.
- JUNGERMANN, HELMUT/PFISTER, HANS-RÜDIGER/FISCHER, KATRIN (1998): Die Psychologie der Entscheidung – Eine Einführung, Heidelberg et al.



- KACHELMEISTER, STEPHEN.J./MESSIER, WILLIAM F. (1990): An Investigation on the Influence of a non-statistical Decision Aid on Auditor Sample Size Decision. In: The Accounting Review, S. 209-226.
- KADOUS, KATHRYN (1996): Discussion of Measuring Cognitive Effort During Analytical Review: A Process-Tracing Framework with Experimental Results. In: Auditing: A Journal of Practice & Theory, Supplement, S. 111-114.
- KAHNEMAN, DANIEL/TVERSKY, AMOS (1982): Variants of Uncertainty. In: Cognition, S. 143-157.
- KAHNEMAN, DANIEL/TVERSKY, AMOS (1984): Choices, Values and Frames. In: American Psychologist, S. 341-350.
- KAHNEMANN, DANIEL/TVERSKY, AMOS (1972): Subjective Probability: A Judgement of Representativeness. In: Cognitive Psychology, S. 430-454.
- KANOUSE, DAVID E. (1972): Language, Labeling and Attribution. In: Jones, E./Kanouse, D./ Kelley, H./Nisbett, R./Valins, S./Weiner, B. (Hrsg.): Attribution: Perceiving the Causes of Behaviour, New York, S. 121-135.
- KAPLAN, CRAIG A./SIMON, HERBERT A. (1990): In Search of Insight. In: Cognitive Psychology, S. 374-419.
- KAPLAN, STEVEN E. (1988): An Examination of the Effect of Presentation Format on Auditors' Expected Value Judgments. In: Accounting Horizons, S. 90-95.
- KAPLAN, STEVEN E./RECKERS, PHILIP M. (1989): An Examination of Information Search During Initial Audit Planning. In: Accounting, Organizations and Society, S. 539-550.
- KAPLAN, STEVEN E./MOECKEL, CINDY/WILLIAMS, JOANNE DEAHL (1992): Auditors' Hypothesis Plausibility Assessments in an Analytical Review Setting. In: Auditing: A Journal of Practice & Theory, Fall, S. 50-65.
- KASK, ALEX W. (1979): Regression and Correlation Analysis. In: The CPA Journal, October, S. 35-41.
- KAYADELEN, ENGIN (2006): Analytische Prüfungen. In: Marten, K.-U./Quick, R./Ruhnke, K. (Hrsg.): Lexikon der Wirtschaftsprüfung, Stuttgart, S. 25-31.
- KELL. WALTER G./BOYNTON, WILLIAM C. (1992): Modern Auditing, 5. Aufl., New York et al.

- KENNEDY, JANE F. (1993): Debiasing Audit Judgment with Accountability: A Framework and Experimental Results. In: *Journal of Accounting Research*, S. 231-245.
- KENNEDY, JANE F. (1995): Debiasing the curse of knowledge in audit judgment. In: *The Accounting Review*, S. 249-273.
- KIDA, THOMAS (1984): The Impact of Hypothesis-Testing Strategies on Auditors' Use of Judgment Data. In: *Journal of Accounting Research*, S. 332-340.
- KIESER, ALFRED/KUBICEK, HERBERT (1978): *Organisationstheorien I – Wissenschaftstheoretische Anforderungen und kritische Analyse klassischer Ansätze*, Stuttgart et al.
- KINNEY, WILLIAM R. JR. (1979): The predictive Power of Limited Information in Preliminary Analytical Review: An Empirical Study. In: *Journal of Accounting Research*, S. 148-165.
- KINNEY, WILLIAM R. JR. (1987): Attention-Directing Analytical Review Using Accounting Ratios: A Case Study. In: *Auditing: A Journal of Practice & Theory*, Spring, S. 59-73.
- KINNEY, WILLIAM R. JR./MCDANIEL, LINDA S. (1996): How to improve effectiveness of substantive analytical procedures. In: *The CPA Journal*, April, S. 52-54.
- KINNEY, WILLIAM R. JR./UECKER, WILFRED C. (1982): Mitigating the consequences of anchoring in auditor judgments. In: *The Accounting Review*, S. 55-69.
- KINNEY, WILLIAM R./BAILEY ANDREW D. (1976): Regression Analysis as a Means of Determining Audit Sample Size: A Comment. In: *The Accounting Review*, S. 396-400.
- KINNEY, WILLIAM R./FELIX, WILLIAM L. (1980): Analytical Review Procedures. In: *Journal of Accountancy*, October, S. 98-103.
- KINNEY, WILLIAM R./HAYNES, CHRISTINE M. (1990): Analytical Procedure Results as Substantive Evidence. In: Srivastava, R. (Hrsg.): *Auditing Symposium X*, Lawrence, S. 83-103.
- KINNEY, WILLIAM R./UECKER, WILFRED C. (1986): *Computer Assisted Analytical Review System*. American Accounting Association, Sarasota.
- KLAGES, ALBRECHT (1968): *Spieltheorie und Wirtschaftsprüfung – Anwendung spieltheoretischer Modelle in der Wirtschaftsprüfung*, Hamburg.

- KLATZKY, ROBERTA.L. (1980): Human Memory: Structures and Processes, San Francisco.
- KLAYMAN, JOSHUA/HA, YOUNG-WON. (1987): Confirmation, disconfirmation and information in hypothesis testing. In: Psychological Review, S. 211-228.
- KLAYMAN, JOSHUA/HA, YOUNG-WON (1989): Hypothesis Testing in Rule Discovery: Strategy, structure and content. In: Journal of Experimental Psychology: Learning, Memory and Cognition, S. 596-604.
- KLITZSCH, WALTER/HELLMUND, UWE/SCHUMANN, KLAUS (1992): Grundlagen der Statistik, Landsberg/Lech.
- KNAPP, CAROL A./KNAPP, MICHAEL C. (2001): The effect of experience and explicit fraud risk assessment in detecting fraud with analytical procedures. In: Accounting, Organizations and Society, S. 25-37.
- KNECHEL, W. ROBERT (1988): The Effectiveness of Nonstatistical Analytical Review Procedures Used as Substantive Audit Tests. In: Auditing: A Journal of Practice & Theory, Fall, S. 87-107.
- KNECHEL, W. ROBERT (1988): The Effectiveness of Statistical Analytical Review as a Substantive Auditing Procedure: A Simulation Analysis. In: The Accounting Review, S. 74-95.
- KNECHEL, W.ROBERT (2001): Auditing : Assurance & Risk, 2. Aufl., Cincinnati et al.
- KÖHLER, RICHARD (1978): Forschungsobjekte und Forschungsstrategien. In: Die Unternehmung, S. 181-196.
- KONRATH, LARRY F. (2002): Auditing: a risk analysis approach, 5. Aufl., Cincinnati et al.
- KOONCE, LISA (1992): Explanation and Counterexplanation During Audit Analytical Review. In: The Accounting Review, S. 59-76.
- KOONCE, LISA (1993): A Cognitive Characterization of Audit Analytical Review. In: Auditing: A Journal of Practice & Theory, Spring, S. 57-108.
- KOONCE, LISA/ANDERSON, URTON/MARCHANT, GARRY (1995): Justification of decisions in auditing. In: Journal of Accounting Research, S. 369-384.

- KOONCE, LISA/PHILLIPS, FRED (1996): Auditor's Comprehension and Evaluation of Client-Suggested Causes in Analytical Procedures. In: Behavioral Research in Accounting, S. 32-49.
- KOROBKIN, B./RUSSEL/ULEN, THOMAS (2000): Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics. In: California Law Review, S. 1053-1147.
- KREIB, JENS-PETER/NEUHAUS, GEORG (2006): Einführung in die Zeitreihenanalyse, Heidelberg.
- KREMS, JOSEF (1996): Expertise und Flexibilität. In: Gruber, H./Ziegler, A. (Hrsg.): Expertiseforschung – Theoretische und methodische Grundlagen, Opladen, S. 80-91.
- KROEBER-RIEL, WERNER/WEINBERG, PETER (1996): Konsumentenverhalten, 6. Aufl., München.
- KROMMES, WERNER (2006): Falsche Aussagen in der Rechnungslegung. In: Zeitschrift Interne Revision, S. 62-73.
- KUBICEK, HERBERT (1977): Heuristische Bezugsrahmen und heuristische angelegte Forschungsdesigns als Elemente einer Konstruktionsstrategie empirischer Forschung. In: Köhler, R. (Hrsg.): Empirische und handlungstheoretische Forschungskonzeptionen in der Betriebswirtschaftslehre, Stuttgart, S. 3-37.
- KÜTING, KARLHEINZ/WEBER, CLAUS-PETER (2006): Der Konzernabschluss - Praxis der Konzernrechnungslegung nach HGB und IFRS, 10. Aufl., Stuttgart.
- KÜTING, KARLHEINZ/WEBER, CLAUS-PETER (1994): Die Bilanzanalyse – Lehrbuch zur Beurteilung von Einzel- und Konzernabschlüssen, Stuttgart.
- LACHNIT, LAURENZ (2004): Bilanzanalyse: Grundlagen – Einzel- und Konzernabschlüsse – Internationale Abschlüsse – Unternehmensbeispiele, 1. Aufl., Wiesbaden.
- LAKATOS, IMRE (1974): Falsifikation und Methodologie wissenschaftlicher Forschungsprogramme. In: Lakatos, I./Musgrave, A. (Hrsg.): Kritik und Erkenntnisfortschritt, Braunschweig, S. 89-189.
- LARCKER, DAVID F./LESSIG, V. PARKER (1983): An Examination of the Linear and Retrospective Process Tracing Approaches to Judgment Modelling, In: The Accounting Review, S. 58-77.

- LARKIN, JILL (1983): The role of problem representation in physics. In: Gentner, D./Stevens, A. (Hrsg.): *Mental Models*, Hillsdale, S. 75-99.
- LAUX, HELMUT (1990): Risiko, Anreiz und Kontrolle – Principal-Agent-Konzept – Einführung und Verbindung mit dem Delegation-Konzept, Heidelberg.
- LAW, SHEE BOON/WILLETT, ROGER (2004): The ability of analytical procedures to signal transaction errors. In: *Managerial Auditing Journal*, S. 869-888.
- LEE, MINWOO/COLBERT, JANET L. (1997): Analytical procedures: management tools for monitoring controls. In: *Management Decision*, S. 392-397.
- LEFFSON, ULRICH (1984): *Bilanzanalyse*, 3. Aufl., Stuttgart.
- LEHNER, FRANZ (1996): Gedanken zur theoretischen Fundierung der Wirtschaftsinformatik und Versuch einer paradigmatischen Einordnung. In: Heilmann, H./Heinrich, L./Roithmayr, F. (Hrsg.): *Information Engineering: Wirtschaftsinformatik im Schnittpunkt von Wirtschafts-, Sozial- und Ingenieurwissenschaften*, München et al., S. 67-85.
- LENZ, HANSRUDI (2002): Prüfungstheorie, verhaltensorientierter Ansatz. In: Ballwieser, W./Coenenberg, A./v.Wysocki, K. (Hrsg.): *Handwörterbuch der Rechnungslegung und Prüfung*, 3. Aufl., Stuttgart, Sp. 1924-1938.
- LESGOLD, ALAN/GLASER, ROBERT/RUBINSON, HARRIET/KLOPFER, DALE/FELTOVICH, PAUL/WANG, YEN (1988): Expertise in a Complex Skill: Diagnosing X-Ray Pictures. In: Michelene, T./Glaser, R./Farr, M. (Hrsg.): *The Nature of expertise*, Hillsdale, S. 311-342.
- LEVIN, IRWIN P./JASPER, J.D./FORBES, WENDY S. (1998): Choosing versus Rejecting options at different stages of Decision making. In: *Journal of Behavioral Decision Making*, S. 193-210.
- LIBBY, PETER (1995): Molecular Bases of the Acute Coronary Syndrome: From Bench to Bedside. In: *Circulation*, S. 2844-2850.
- LIBBY, ROBERT (1981): *Accounting and Human Information Processing: Theory and Applications*, Englewood Cliffs.
- LIBBY, ROBERT (1985): Availability and Generation of Hypotheses in Analytical Review. In: *Journal of Accounting Research*, S. 648-667.
- LIBBY, ROBERT/FREDERICK, DAVID M. (1990): Experience and the Ability to Explain Audit Findings. In: *Journal of Accounting Research*, S. 348-367.

- LIN, KENNY Z. /FRASER, IAN A. M./HATHERLY, DAVID J. (2003): Auditor analytical review judgment: a performance evaluation. In: *The British Accounting Review*, S. 19-34.
- LIN, KENNY Z./FRASER, IAN A. M./HATHERLY, DAVID J. (2000): An Experimental Study of Auditor Analytical Review Judgements. In: *Journal of Business Finance & Accounting*, S. 821-857.
- LIPE, MARLYS GASCHO (1997): Commentary on The Effect of Justification, Task Complexity and Experience/Training on Problem-Solving Performance. In: *Behavioral Research in Accounting*, S. 117-122.
- LOEBBECKE, JAMES K./STEINBERT, PAUL J. (1987): An Investigation of the Use of Preliminary Analytical Review to Provide Substantive Audit Evidence. In: *Auditing: A Journal of Practice & Theory*, Spring, S. 74-89.
- LOITLSBERGER, ERICH (1968): Die Buchführung als spieltheoretisches Problem. In: *Der Österreichische Betriebswirt*, S. 137-179.
- LOREK, KENNETH S./BRANSON, BRUCE C./ICERMAN, RHODA C. (1992): On the Use of Time-Series Models as Analytical Procedures. In: *Auditing: A Journal of Practice & Theory*, Fall, S. 66-87.
- LÜCK, WOLFGANG (1991): *Wirtschaftsprüfung und Treuhandwesen – Institutionelle und funktionale Aspekte der Betriebswirtschaftlichen Prüfungslehre*, 2. Aufl., Stuttgart.
- LÜCK, WOLFGANG (1998): *Lexikon der Rechnungslegung und Abschlussprüfung*, 4. Aufl., München et al.
- MACK, WOLFGANG (1996): Expertise und Intelligenz. In: Gruber, H./Ziegler, A. (Hrsg.): *Expertiseforschung – Theoretische und methodische Grundlagen*, Opladen, S. 92-114.
- MAHATHEVAN, PREMILA (1997): Auditors' Use and Perception of Analytical Procedures: Evidence from Singapore. In: *International Journal of Auditing*, S. 225-239.
- MAJOR, BRENDA (1980): Information Acquisition and Attribution Processes. In: *Journal of Personality and Social Psychology*, S. 1010-1023.
- MARTEN, KAI-UWE/QUICK, REINER/RUHNKE, KLAUS (2003): *Wirtschaftsprüfung. Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen*, 2. Aufl., Stuttgart.

- MARTEN, KAI-UWE/QUICK, REINER/RUHNKE, KLAUS (2007): Wirtschaftsprüfung. Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen, 3. Aufl., Stuttgart.
- MARTEN, KAI-UWE/QUICK, REINER/RUHNKE, KLAUS (2006): Lexikon der Wirtschaftsprüfung. Nach nationalen und internationalen Normen, Stuttgart.
- MATSUMURA, ELLA M./TUCKER, ROBERT R. (1992): Fraud Detection: A Theoretical Foundation. In: The Accounting Review, S. 753-782.
- MCDANIEL, LINDA (1990): The effect of time pressure and audit program structure on audit performance. In: Journal of Accounting Research, S. 267-285.
- MCDANIEL, LINDA S./KINNEY, WILLIAM R. JR. (1995): Expectation-Formation Guidance in the Auditor's Review of Interim Financial Information. In: Journal of Accounting Research, S. 59-76.
- McKEE, THOMAS E. (1982): Developments in Analytical Review. In: The CPA Journal, January, S. 36-42.
- McKEE, THOMAS E. (1989): Modern Analytical Auditing, New York.
- McKEITHEN, KATHRINE B./REITMAN, JUDITH S./RUETER, HENRY H./HIRTLE, STEPHEN C. (1981): Knowledge organization and skill differences in computer programmers. In: Cognitive Psychology, S. 307-325.
- MERTZ, HERBERT P. (1963): Prüfschärfe bei Revisionen – Ein Beispiel zur Entscheidungstheorie. In: Ablauf- und Planungsforschung, S. 315-319.
- MERVIS, CAROLYN B./ROSCH, ELEANOR (1981): Categorization of Natural Objects. In: Annual Review of Psychology, S. 89-115.
- MEYER, DAVID E./SCHVANEVELDT, ROGER W. (1976): Meaning, Memory Structures and Mental Processes. In: Science, S. 27-33.
- MIGRINI, MARK J./MITTERMAIER, LINDA J. (1997): The Use of Benford's Law as an Aid in Analytical Procedures. In: Auditing: A Journal of Practice & Theory, Fall, S. 52-67.
- MILLS, TINA Y. (1996): The Effect of Cognitive style on External Auditors' Reliance Decisions on Internal Audit Functions. In: Behavioural Research in Accounting, S. 49-73.
- MORTON, SANFORD (1993): Strategic Auditing for Fraud. In: The Accounting Review, S. 825-839.

- MUELLER, JENNIFER M./ANDERSON, JOHN C. (2002): Decision Aids for Generating Analytical Review Alternatives: The Impact of Goal Framing and Audit-Risk Level. In: Behavioural Research in Accounting, S. 158-177.
- MÜLLER, CHRISTIAN (1996): Entwicklung eines wissensbasierten Systems zur Unterstützung analytischer Prüfungshandlungen im Rahmen der Jahresabschlussprüfung, Frankfurt am Main.
- MÜLLER, CHRISTIAN/KROPP, MANFRED (1992): Die Überprüfung der Plausibilität von Jahresabschlüssen. In: Der Betrieb, S. 149-158.
- MULLIGAN, CHRISTINA/INKSTER, NICOLA (1999): The Use of Analytical Procedures in the United Kingdom. In: International Journal of Auditing, S. 107-120.
- NELSON, MARK W. (1993): The Effects of Error Frequency and Accounting Knowledge on Error Diagnosis in Analytical Review. In: The Accounting Review, S. 804-824.
- NELSON, MARC W. (1994): The Learning and Application of Frequency Knowledge in Audit Judgment. In: Journal of Accounting Literature, S. 185-211.
- NEWBOLD, PAUL/BOS, THEODORE (1990): Introductory Business Forecasting, Cincinnati.
- NEWELL, ALLEN/SIMOM, HERBERT A. (1972): Human Problem Solving, Englewoods Cliffs (NJ).
- NG, TERENCE BU-PEOW/GREEN, WENDY/SIMNETT, ROGER (2001): The Effects of Fraud Risk and Management Representation on Auditors' Hypothesis Generation. In: ABACUS, S. 352-368.
- NICKERSON, RAYMOND S. (1984): Retrieval Inhibition from Part-set Cueing: A Persisting Enigma in Memory Research. In: Memory and Cognition, S. 531-552.
- NOVIK, LAURA R. (1988): Analogical Transfer, Problem Similarity and Expertise. In: Journal of Experimental Psychology: Learning, Memory and Cognition, S. 510-520.
- O'CLOCK, PRISCILLA/DEVINE, KEVIN (1995): An Investigation of Framing and Firm Size on the Auditor's Going Concern Decision. In: Accounting and Business Research, S. 197-207.
- O'DONNELL, ED (2004): Use forward versus backward reasoning during audit analytical procedures: evidence from a computerised-process-tracing field study. In: Accounting and Finance, S. 75-95.



- O'DONNELL, ED/LEHMAN, MARK (1999): Information processing strategy during analytical procedures: where and how auditors focus their attention. In: *Accounting Forum*, S. 241-273.
- O'DONNELL, ED (1996): Measuring Cognitive Effort During Analytical Review: A Process-Tracing Framework with Experimental Results (Reply). In: *Auditing: A Journal of Practice & Theory*, Supplement, S. 118.
- O'DONNELL, ED (2002): Evidence of an Association between Error-Specific Experience and Auditor Performance during Analytical Procedures. In: *Behavioural Research in Accounting*, S. 179-195.
- O'DONNELL, ED/SCHULTZ, JR., JOSEPH J. (2003): The Influence of business-process-focused audit support software on analytical procedures judgments. In: *Auditing: A Journal of Practice & Theory*, September, S. 265-279.
- OPWIS, KLAUS (1992): *Kognitive Modellierung: Zur Verwendung wissensbasierter Systeme in der psychologischen Theoriebildung*, Bern.
- OPWIS, KLAUS/LÜER, GERD (1996): Modelle der Repräsentation von Wissen. In: Albert, D./Stapf, K.-H.. (Hrsg.): *Enzyklopädie der Psychologie. Gedächtnispsychologie: Erwerb, Nutzung und Speicherung von Informationen*, Göttingen, S. 337-431.
- PATEL, VIMLA L./GROEN, GUY J. (1986): Knowledge-based solutions strategies in medical reasoning. In: *Cognitive Science*, S. 91-116.
- PEECHER, MARK E. (1996): The Influence of Auditors' Justification Processes on Their Decisions: A Cognitive Model and Experimental Evidence. In: *Journal of Accounting Research*, S. 125-140.
- PEEMÖLLER, VOLKER H. (1992): Stichwort "Direkte und Indirekte Prüfung". In: Coenenberg A./v.Wysocki, K. (Hrsg.): *Handwörterbuch der Revision*, 2. Aufl., Stuttgart, Sp. 343-348.
- PENNINGTON, NANCY/HASTIE, REID (1992): Explaining the Evidence: Tests of the Story Model for Juror Decision Making. In: *Journal of Personality and Social Psychology*, S. 189-206.
- PENNINGTON, NANCY/HASTIE, REID (1988): Explanation-based Decision Making: The Effect of Memory Structure on Judgment. In: *Journal of Experimental Psychology: Learning, Memory and Cognition*, S. 521-533.

- PENNY, JULIA (2003): Do the numbers make sense? In: Accountancy, November, S. 92-93.
- PETERSON, BONITA K./WONG-ON-WING, BERNARD (2000): An Examination of the Positive Test Strategy in Auditors' Hypothesis Testing. In: Behavioral Research in Accounting, S. 257-277.
- PINCUS, KAREN (1990): Auditor Individual Differences and Fairness of Presentation Judgment. In: Auditing: A Journal of Practice & Theory, Fall, S. 150-166.
- PINCUS, KAREN. (1989): The Efficacy of a Red Flag Questionnaire for Assessing the Possibility of Fraud. In: Accounting, Organizations and Society, S. 153-163.
- PINNEKAMP, HEINZ-JÜRGEN/SIEGMANN, FRANK (1993): Deskriptive Statistik, 3. Aufl., München et al.
- POPPER, KARL (1994): Logik der Forschung, 10. Aufl., Tübingen.
- PORTER, BRENDA/SIMON, JON/HATHERLY, DAVID (2003): Principles of External Auditing, 2. Aufl., West Sussex.
- PUTZ-OSTERLOH, WIEBKE (1983): Über Determinanten komplexer Problemlösungen und Möglichkeiten zu ihrer Erfassung. In: Sprache und Kognition, S. 100-116.
- QUICK, REINER (1999): Prüfungsmethoden im Spiegel der Forschung. In: Richter, M. (Hrsg.): Theorie und Praxis der Wirtschaftsprüfung II, Wirtschaftsprüfung und ökonomische Theorie - Prüfungsmarkt - Prüfungsmethoden - Urteilsbildung, 2. Symposium der KPMG/Universität Potsdam zur Theorie und Praxis der Wirtschaftsprüfung am 9. und 10. Oktober 1998 in Potsdam, Berlin, S. 177-234.
- QUICK, REINER (2002): Plausibilitätsbeurteilungen. In: Ballwieser, W./Coenenberg, A./v.Wysocki, K. (Hrsg.): Handwörterbuch der Rechnungslegung und Prüfung, 3. Aufl., Stuttgart, Sp. 1686-1693.
- QUICK, REINER (2004): Externe Pflichtrotation: eine adäquate Maßnahme zur Stärkung der Unabhängigkeit des Abschlussprüfers? In: Die Betriebswirtschaft, S. 487-508.
- QUICK, REINER (2006): Prüfung, Beratung und Unabhängigkeit des Abschlußprüfers - eine Analyse der neuen Unabhängigkeitsnormen des HGB im Lichte empirischer Forschungsergebnisse. In: Betriebswirtschaftliche Forschung und Praxis, S. 42-61.
- QUICK, REINER/ WOLZ, MATTHIAS (2007): Bilanzierung in Fällen. Grundlagen, Aufgaben und Lösungen nach HGB und IFRS, 3. Aufl., Stuttgart.

- RANDOLPH, W. ALAN/ELLOY, DAVID F. (1989): How Can OD Consultants and Researchers Assess Gamma Change? A Comparison of Two Analytical Procedures. In: Journal of Management, S. 633-648.
- REIMANN, PETER (1998): Unterstützung kollaborativer Arbeitsformen in Teleteaching-Szenarien.
- REIMERS, JANE L./FENNEMA, M. G. (1999): The Audit Review Process and Sensitivity to Information Source Objectivity. In: Auditing: A Journal of Practice & Theory, Spring, S. 117-123.
- REIMERS, JANE L./BUTLER, STEPHEN A. (1992): The Effect of Outcome Knowledge on Auditors' Judgemental Evaluations. In: Accounting, Organizations and Society, S. 185-194.
- RICHTER, MARTIN (1997a): Entwicklungslinien, Nutzen und Grenzen der Forschung im Prüfungswesen. In: Richter, M. (Hrsg.): Theorie und Praxis der Wirtschaftsprüfung – Abschlussprüfung, Interne Revision, Kommunale Rechnungsprüfung, Berlin, S. 249-300.
- RICHTER, MARTIN (1997b): Empirische Untersuchungen in der deutschsprachigen Prüfungslehre. In: Richter, M (Hrsg.): Theorie und Praxis der Wirtschaftsprüfung – Abschlussprüfung, Interne Revision, Kommunale Rechnungsprüfung, Berlin, S. 249-300.
- RICHTER, MARTIN (1999): Konzeptioneller Bezugsrahmen für eine realwissenschaftliche Theorie betriebswirtschaftlicher Prüfungen. In: Richter, M. (Hrsg.): Theorie und Praxis der Wirtschaftsprüfung II, Wirtschaftsprüfung und ökonomische Theorie – Prüfungsmarkt – Prüfungsmethoden – Urteilsbildung, Berlin, S. 263-307.
- RIECK, CHRISTIAN (1993) Spieltheorie, Wiesbaden.
- ROBINSON, LORI B./HASTIE, REID (1985): Revision of beliefs: When a hypothesis is eliminated from consideration. In: Journal of Experimental Psychology: Human Perception and Performance, S. 443-456.
- ROMHARDT, KAI (2001): Die Organisation aus der Wissensperspektive – Möglichkeiten und Grenzen der Intervention, Dissertation, Hamburg.
- RÖNZ, BERND/FÖRSTER, ERHARD (1992): Regressions- und Korrelationsanalyse, Wiesbaden.
- ROSPLOCK, MICHAEL F. (2001): Advanced Analytical Techniques for Performing Forensic Financial Analysis. In: Business Credit, Vol. 103, Heft 6, S. 26-31.

- ROTHE, HEINZ-JÜRGEN/SCHINDLER, MARION (1996): Expertise und Wissen. In: Gruber, H./Ziegler, A. (Hrsg.): Expertiseforschung – Theoretische und methodische Grundlagen, Opladen, S. 35-57.
- RUBINSTEIN, SERGEJ L. (1972): Das Denken und die Wege seiner Erforschung, 4. Aufl., Berlin (Ost).
- RUHNKE, KLAUS (2000): Normierung der Abschlussprüfung, Stuttgart.
- RUHNKE, KLAUS (2006): Prüfungstheorie. In: Marten, K.-U./Quick, R./Ruhnke, K. (Hrsg.): Lexikon der Wirtschaftsprüfung, S. 650-655.
- RUHNKE, KLAUS/LUBITZSCH, KAY (2006): Abschlussprüfung und das neue Aussagenkonzept der IFAC: Darstellung, Beweggründe und Beurteilung. In: Die Wirtschaftsprüfung, S. 366-375.
- RUSSO, JOSEPH A. (1994): An Investigation of Auditor Problem-Solving Behavior in an Unfamiliar Task Situation, Dissertation, State University of New Jersey, Newark.
- SCHANZ, GÜNTHER (1988): Methodologie für Betriebswirte, 2. Aufl., Stuttgart.
- SCHANZ, GÜNTHER (1993): Verhaltenswissenschaften und Betriebswirtschaftslehre. In: Wittmann, W./Kern, W./Köhler, R./Küpper, U./v.Wysocki, K. (Hrsg.): HWB der Betriebswirtschaft, Band 2, 5. Aufl., Stuttgart, Sp. 4521-4532.
- SCHANZ, GÜNTER (1978): Verhalten in Wirtschaftsorganisationen, München.
- SCHANZ, GÜNTHER (1979): Betriebswirtschaftslehre als Sozialwissenschaft, Stuttgart et al.
- SCHANZ, GÜNTHER (1988): Methodologie für Betriebswirte, Stuttgart.
- SCHLITTGEN, RAINER (2003): Einführung in die Statistik – Analyse und Modellierung von Daten, 10. Aufl., München.
- SCHMIDT, MARTIN (2006): Abschlusss Aussagen. In: Marten, K.-U./Quick, R./Ruhnke, K. (Hrsg.): Lexikon der Wirtschaftsprüfung, S. 3-8.
- SCHNEIDER, DIETER (1981): Geschichte betriebswirtschaftlicher Theorie, München et al.
- SCHNEIDER, DIETER. (1995) Betriebswirtschaftslehre: Grundlagen (Band 1), 2. Aufl., München.
- SCHNEIDER, DIETER (1984) Dilettantismusgefahr und „grenzüberschreitendes wissenschaftliches Arbeiten“ – Anmerkungen zu einem Beitrag von Henning Egner. In: Zeitschrift für betriebswirtschaftliche Forschung, S. 1070-1073.

- SCHNELL, RAINER/HILL, PAUL B./ESSER, ELKE (2005): Methoden der empirischen Sozialforschung. 7. Aufl., München et al.
- SCHORB, MANFRED (1994): Verhaltensorientiertes FuE-Controlling – Ein ganzheitliches Konzept für ein mittleres Industrieunternehmen, München.
- SCHREIBER, STEFAN M. (2000): Das Informationsverhalten von Wirtschaftsprüfern: Eine Prozessanalyse aus verhaltenswissenschaftlicher Perspektive, Wiesbaden.
- SCHULTZ, JR./JOSEPH J. (1994): Commentary on The Functionality of Decision Heuristic: Reliance on Prior Audit Adjustments in Evidential Planning. In: Behavioral Research in Accounting, S. 90-96.
- SCHÜTTE, REINER (1999): Basispositionen in der Wirtschaftsinformatik. In: Becker, J./König, W./Schütte, R./Wendt, O./ Zelewski, S. (Hrsg.): Wirtschaftsinformatik und Wissenschafts-theorie. Bestandsaufnahme und Perspektiven, Wiesbaden, S. 211-242.
- SCOTT, DAVID A./WALLACE, WANDA A.(1994): A second look at an old tool: Analytical procedures. In: The CPA Journal, March, S. 30-35.
- SCOTT, DAVID A./WALLACE, WANDA A.(1994): Analytical procedures. In: The CPA Journal, December, S. 64-65.
- SEIFFERT, HELMUT (1996): Einführung in die Wissenschaftstheorie 1, 12. Aufl., München.
- SEIPEL, CHRISTIAN/RIEKER, PETER (2003): Integrierte Sozialforschung. Konzepte und Methoden der qualitativen und quantitativen empirischen Forschung, Weinheim et al.
- SHAKLEE, HARRIET/FISCHHOFF, BARUCH (1982): Strategies of Information Search in Causal Analysis. In: Memory and Cognition, S. 520-530.
- SHANTEAU, JAMES (1989): Cognitive Heuristics and Biases in Behavioural Auditing: Review, Comments and Observation. In: Accounting, Organizations and Society, S. 165-177.
- SHANTEAU, JAMES (1993): Discussion of „Expertise in Auditing“. In: Auditing: A Journal of Practice & Theory, Supplement, S. 51-56.
- SHERMAN, STEVEN J./CIALDINI, ROBERT B./SCHWARTZMAN, DONNA F./REYNOLDS, KIM D. (1985): Imaging can Heighten or Lower the Perceived Likelihood of Contracting a disease: The mediating Effect of Ease of Imagery. In: Personality and Psychology Bulletin, S. 118-127.

- SHERMAN, STEVEN J./MCMULLEN, MATTHEW/GAVANSKI, IGOR (1992): Natural sample spaces and the inversion of conditional judgments. In: *Journal of Experimental Social Psychology*, S. 401-421.
- SHIELDS, MICHAEL D./SOLOMON, IRA/WALLER, WILLIAM S. (1998): Auditors' usage of unaudited book values when making presampling audit value estimates. In: *Contemporary Accounting Research*, S. 1-18.
- SIEBEN, GÜNTHER (1998): Zum 125. Geburtstag von Eugen Schmalenbach, Betriebswirtschaftslehre als angewandte Wissenschaft. In: *Handelsblatt* vom 7.5.1998, S. 5.
- SIEGEL, SIDNEY (1997): *Nichtparametrische statistische Methoden*, 4. Aufl., Eschborn.
- SLOVIC, PAUL/KUNREUTHER, HOWARD/WHITE, GILBERT F. (1974): Decision processes, rationality, and adjustment to natural hazards. In: White G. (Hrsg.): *Natural hazards. Local, national and global*, New York, S. 187-205.
- SMITH, DALE GEORGE (1989): *Analytical Review*, Toronto.
- SMITH, GEOFFREY/PSAROS, JIM/HOLMES, SCOTT (1999): A Research Note on the Use and Perceived Usefulness of Analytical Procedures by Australian Auditors. In: *Australian Accounting Review*, Vol. 9, No. 2, S. 64-72.
- SMITH, GERALD F. (1988): Toward a Heuristic Theory of Problem Structuring. In: *Management Science*, S. 1489-1506.
- SMITH, JANE F./KIDA, THOMAS (1991): Heuristics and Biases: Expertise and task realism in auditing. In: *Personality and Psychological Bulletin*, S. 472-489.
- SOLOMON, IRA/SHIELDS, MICHAEL D. (1995): Judgment and Decision-Making Research in Auditing. In: Ashton, R./Ashton, A. (Hrsg.): *Judgment and Decision-Making Research in Accounting and Auditing*, New York, S. 137-175.
- SOLOMON, IRA/SHIELDS, MICHAEL D./WHITTINGTON, O. RAY (1999): What do industry-specialist auditors know? In: *Journal of Accounting Research*, S. 191-207.
- SPERL, ANDREAS (1978): *Prüfungsplanung*, Düsseldorf.
- SPRINKLE, GEOFFREY B./TUBBS, RICHARD M. (1998): The Effects of Audit Risk and Information Importance on Auditor Memory During Working Paper Review. In: *The Accounting Review*, S. 475-502.
- SRIVASTAVA, RAJENDRA P./WRIGHT, ARNOLD/MOCK, THEODORE J. (2002): Multiple hypothesis evaluation in auditing. In: *Accounting and Finance*, S. 251-277.

- STÄHLIN, WIGAND (1973) Theoretische und technologische Forschung in der Betriebswirtschaftslehre.
- STEINER, I. (1972): Group Processes and Productivity, New York.
- STELZL, IVAN D. (1995): Experiment. In: Roth, E. (Hrsg.): Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis, 4. Aufl., München et al., S. 108-125.
- STEWART, TREVOR (1978): More Effective Reviews with Regression Analyses. In: CA-Magazine, May, S. 68-72.
- STIER, WINFRIED (1996): Empirische Forschungsmethoden, Berlin.
- STÖKER, ELISABETH (1987): Einführung in die Wissenschaftstheorie, 3. Aufl., Darmstadt.
- STRUBE, GERHARD/SCHLIEDER, CHRISTOPH (1998): Wissensrepräsentation im Symbolverarbeitungsansatz. In: Klix, F./Spada, H. (Hrsg.): Enzyklopädie der Psychologie: Wissen, Vol. 6, Göttingen, S. 501-530.
- TABOR, RICHARD H./WILLIS, JAMES T. (1985): Empirical Evidence on the Changing Role of Analytical Review Procedures. In: Auditing: A Journal of Practice & Theory, Spring, S. 93-109.
- TAN, HUN-TONG (1995): Effects of Expectations, Prior Involvement, and Review Awareness on Memory for Audit Evidence and Judgment. In: Journal of Accounting Research, S. 113-135.
- TAN, HUN-TONG/KAO, ALISON (1999): Accountability Effects on Auditors' Performance: The Influence of Knowledge, Problem Solving Ability and Task Complexity. In: Journal of Accounting Research, S. 209-223.
- TAN, HUN-TONG/NG, TERENCE BU-PEOW/MAK, BOBBY WAI-YEONG (2002): The Effect of Task Complexity on Auditors' Performance: The Impact of Accountability and Knowledge. In: Auditing: A Journal of Practice & Theory, September, S. 81-94.
- TAPLIN, JULIAN R. (1975): Evaluation of Hypotheses in Concept Identification. In: Memory and Cognition, S. 85-96.
- TERLINDE, CHRISTIAN (2005): Aufdeckung von Bilanzmanipulationen in der deutschen Prüfungspraxis, Wiesbaden.
- TEIGEN, KARL (1983): Studies in Subjective Probability III: The Unimportance of Alternatives. In: Scandinavian Journal of Psychology, S. 97-105.

- TEIGEN, KARL (1988): When are low probability judged to be probable? Effects of outcome-set characteristics on verbal probability estimates. In: *Acta Psychologica*, S. 157-174.
- TROTMAN, KEN T./SNG, JENNIFER (1989): The Effect of Hypothesis Framing, Prior Expectations and Cue Diagnosticity and Auditors' Information Choice. In: *Accounting, Organizations and Society*, S. 565-576.
- TUBBS, RICHARD M. (1992): The Effect of Experience on the Auditor's Organization and Amount of Knowledge. In: *The Accounting Review*, S. 783-801.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1972): Subjective Probability: A Judgment of Representativeness. In: *Cognitive Psychology*, S. 430-454.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1973): On the Psychology of Prediction. In: *Psychological Review*, S. 237-251.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1974): Judgement under Uncertainty: Heuristics and Biases. In: *Science*, S. 1124-1131.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1979): Prospect Theory: An Analysis of Decision under Risk. In: *Econometrica*, S. 263-292.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1981): The Framing of Decisions and the Rationality of Choice. In: *Science*, S. 453-458.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1982): Judgments of and by Representativeness. In: Kahneman, D./Slovic, P./Tversky, A. (Hrsg.): *Judgment under Uncertainty: Heuristics and Biases*, Cambridge et al., S. 84-98.
- TVERSKY, AMOS/KAHNEMAN, DANIEL (1986): Rational Choice and the framing of Decisions. In: *Journal of Business*, S. 251-278.
- ULRICH, HANS (1982): Anwendungsorientierte Wissenschaft. In: *Die Unternehmung*, S. 1-10.
- ULRICH, PETER/FLURI, EDGAR (1992): *Management: Eine konzentrierte Einführung*, 6. Aufl, Bern et al.
- VAN WALLENDael, LORI/HASTIE, REID (1990): Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. In: *Memory and Cognition*, S. 240-250.
- VOß, WILHELM (1926): Berufspolitisches und Fachliches vom Verbandstag in Köln a. Rh. In: *Zeitschrift des Verbandes Deutscher Bücherrevisoren*, Leipzig, S. 169-179.



- WALLACE, WANDA A. (1983a): Analytical Review: Misconceptions, Applications and Experience – Part I. In: The CPA Journal, January, S. 24-73.
- WALLACE, WANDA A. (1983b): Analytical Review: Misconceptions, Applications and Experience – Part II. In: The CPA Journal, February, S. 18-27.
- WALLACE, WANDA A. (1991): Auditing, 2. Aufl., Englewood Cliffs.
- WEBER, RON (1980): Some Characteristics of the Free Recall of Computer Controls by EDP Auditors. In: Journal of Accounting Research, S. 214-239.
- WEIBLER, JÜRGEN (1996): Ökonomische vs. Verhaltenswissenschaftliche Ausrichtung der Personalwirtschaftslehre – Eine notwendige Kontroverse? In: Der Betriebswirt, S. 649-665.
- WHEELER, STEPHEN/PANY, KURT (1990): Assessing the Performance of Analytical Procedures: A Best Case Scenario. In: The Accounting Review, S. 557-577.
- WILD, JOHN J./BIGGS, STANLEY F. (1990): Strategic Considerations for Unaudited Account Values in Analytical Review. In: The Accounting Review, S. 227-241.
- WILSON, ARLETTE C. (1991): Use of Regression Models as Analytical Procedures: An Empirical Investigation of Effect of Data Dispersion on Auditor Decisions: In: Journal of Accounting, Auditing & Finance, S. 365-381.
- WILSON, ARLETTE C./COLBERT, JANE L. (1991): Analytical Procedures as Substantive Tests. In: The National Public Accountant, May, S. 38-41.
- WILSON, ARLETTE C./COLBERT, JANET (1989): An Analysis of Simple and Rigorous Decision Models as Analytical Procedures. In: Accounting Horizons, S. 79-83.
- WILSON, ARLETTE C./GLEZEN, WILLIAM G. (1989): Regression Analysis in Auditing: A Comparison of Alternative Investigation Rules - Some Further Evidence. In: Auditing: A Journal of Practice & Theory, Spring, S. 90-100.
- WIPPICH, WERNER (1984): Lehrbuch der angewandten Gedächtnispsychologie, Band 1, Stuttgart et al.
- WITTE, EBERHARD (1985): Der Wirtschaftsprüfer als empirischer Forscher. In: Gross, G. (Hrsg.): Der Wirtschaftsprüfer im Schnittpunkt nationaler und internationaler Entwicklungen, Düsseldorf, S. 177-186.
- WÖHE GÜNTER (1959): Methodologische Grundprobleme der Betriebswirtschaftslehre, Meisheim.

- WOLZ, MATTHIAS (2003): Wesentlichkeit im Rahmen der Jahresabschlussprüfung, Düsseldorf.
- WRIGHT, ARNOLD M./BEDARD, JEAN C. (2000): Decision Processes in Audit Evidential Planning: A Multistage Investigation. In: Auditing: A Journal of Practice & Theory, Spring, S. 123-143.
- WRIGHT, ARNOLD/ASHTON, ROBERT H. (1989): Identifying Audit Adjustments with Attention-Directing Procedures. In: The Accounting Review, S. 710-728.
- WRIGHT, SALLY/WRIGHT, ARNOLD M. (1997): The Effect of Industry Experience on Hypothesis Generation and Audit Planning Decisions: In: Behavioural Research in Accounting, S. 273-294.
- WRIGHT, WILLIAM. F. (1982): Comparison of the Lens and Subjective Probability Paradigms for Financial Research Purposes. In: Accounting, Organizations and Society, S. 65-75.
- WYSOCKI, KLAUS VON (1993): Prüfung und Kontrolle: Empirische Befunde im deutschen Prüfungswesen. In: Hauschild, J./Grün, O. (Hrsg.): Ergebnisse empirischer betriebswirtschaftlicher Forschung – Zu einer Realtheorie der Unternehmung, Festschrift für Eberhard Witte, Stuttgart, S. 909-923.
- WYSOCKI, KLAUS VON (2002): Prüfungstheorie, messtheoretischer Ansatz. In: Ballwieser, W./Coenenberg, A./v.Wysocki, K. (Hrsg.): Handwörterbuch der Rechnungslegung und Prüfung, 3. Aufl., Stuttgart, Sp. 1886-1899.
- WYSOCKI, KLAUS VON (2003): Wirtschaftliches Prüfungswesen, Band 3, München.
- YANIV, ILAN/SCHUL, YAACOV (1997): Elimination and Inclusion Procedures in Judgment. In: Journal of Behavioral Decision Making, S. 211-220.
- YANIV, ILAN/SCHUL, YAACOV (2000): Acceptance and Elimination Procedures in Choice: Noncomplementarity and the Role of implied Status Quo. In: Organizational Behaviour and Human Decision Processes, S. 293-313.
- YIP-OW, JACKSON/TAN, HUN-TONG (2000): Effects of the preparer's justification on the reviewer's hypothesis generation and judgment in analytical procedures. In: Accounting, Organizations and Society, S. 203-215.
- ZIMBARDO, PHILIP G. (1995): Psychologie, 6. Aufl., Berlin et al.
- ZÜND, ANDRÉ (1982): Revisionslehre, Zürich.

## Anhang 1: Text der Sammelmail

Liebe Kollegen,

Herr Dipl.-Kfm. Engin Kayadelen, Mitarbeiter am Lehrstuhl für Rechnungswesen, Controlling und Wirtschaftsprüfung der Technischen Universität Darmstadt, sucht im Rahmen eines Forschungsprojektes Teilnehmer für die Durchführung einer empirischen Studie zum Thema „analytische Prüfungen“. Die Bearbeitung einer diesbezüglichen Fallstudie erfolgt online und kann von den Teilnehmern von jedem Ort aus und zu jeder gewünschten Zeit vorgenommen werden. Mir wurde zugesichert, dass sich die Bearbeitungszeit in überschaubaren Grenzen hält (etwa 25 – 30 Minuten) und absolut anonym verläuft.

Da unsere Niederlassung enge Beziehungen zum Lehrstuhlinhaber, Herrn Prof. Dr. Reiner Quick, pflegt, haben wir uns bereit erklärt, dieses Projekt zu unterstützen und den Kontakt zu potenziellen Teilnehmern herzustellen. Daher bitte ich alle interessierten Mitarbeiter, die Fallstudie bis zum **23. Juni 2006** online zu bearbeiten. Als Dankeschön werden unter allen Teilnehmern ein **Apple iPod** im Wert von 250 € und vier **Amazon-Buchgutscheine** zu jeweils 25 € verlost. Nähere Informationen können Sie dem beigefügten Anhang entnehmen.

Mit freundlichen Grüßen

xxx

## Anhang 2: Text des Anschreibens

Sehr geehrte Damen und Herren,

im Rahmen eines Forschungsprojektes führe ich eine Studie zur Anwendung von analytischen Prüfungen durch. Ihre Teilnahme ist für den Erfolg der Untersuchung von großer Bedeutung, so dass ich Sie herzlich bitten möchte, die Fallstudie zu bearbeiten.

Ziel dieser Studie ist es, Informationen zur Durchführung von analytischen Prüfungen im Rahmen der Jahresabschlussprüfung zu erhalten. Weiterhin möchte ich ermitteln, welche Faktoren sich positiv auf die Effektivität und Effizienz von analytischen Prüfungen auswirken. Sicherlich werden die gewonnenen Erkenntnisse auch für die Praxis von hohem Nutzen sein.

Ich wäre Ihnen sehr verbunden, wenn Sie an der fallstudiengestützten Studie mitwirken würden. Selbstverständlich werden Ihre Angaben streng vertraulich und anonym behandelt. Bitte bearbeiten Sie die Fallstudie innerhalb der nächsten **vier** Wochen. Der eigens hierzu erstellte Internetauftritt steht Ihnen bis zum **23. Juni 2006** online unter der Adresse

<http://www.bwl.tu-darmstadt.de/bwl4/fallstudie/anfang.html>

zur Verfügung.

Insgesamt werden Sie zur Bearbeitung der Fallstudie etwa 25 – 30 Minuten benötigen. Als Dankeschön würde ich gerne unter allen Teilnehmern einen **Apple iPod im Wert von 250 €** und **vier Amazon-Buchgutscheine zu jeweils 25 €** verlosen. Hierzu brauchen Sie mir nur eine unkommentierte E-Mail an die Adresse zu schicken, die Ihnen zum Ende der Fallstudienbearbeitung angezeigt wird. Sie können davon ausgehen, dass der Versand dieser E-Mail nicht mit den Daten aus der Fallstudienbearbeitung in Verbindung gebracht werden kann.

Gleichzeitig werde ich allen Absendern der E-Mail die Untersuchungsergebnisse sowie die Lösung der Fallstudie zukommen lassen, sobald die Auswertung der Ergebnisse abgeschlossen wurde. Die Auswertung erfolgt in aggregierter und anonymisierter Form, so dass keine Rückschlüsse auf einzelne Personen möglich sind.

Sollten Sie Fragen zur Studie im Allgemeinen oder zur Vorgehensweise bei der Fallbearbeitung im Speziellen haben, können Sie mich gerne unter den oben aufgeführten Kontaktmöglichkeiten erreichen.

Für Ihre Teilnahme an dieser Untersuchung darf ich mich schon jetzt recht herzlich bedanken.

Mit freundlichen Grüßen

Engin Kayadelen

### Anhang 3: Liste der 19 möglichen Tests

1. Vergleiche das Produktsortiment nach Monaten mit dem Vorjahr (2,5 Stunden).
2. Hole vergleichende Jahresabschlussdaten von Mandanten der gleichen Branche ein und stelle fest, ob sich deren Gewinnmargen ebenfalls erhöht haben (3 Stunden).
3. Überprüfe die Umsatzabgrenzung zum Abschlussstichtag, um die Korrektheit der Periodenzuordnung der Umsätze festzustellen (3 Stunden).
4. Überprüfe, ob alle Umsätze des abgelaufenen Geschäftsjahres auch zu entsprechenden Einbuchungen der Umsatzkosten geführt haben, durch Kontrolle, ob die letzten Eintragungen in das Warenausgangsbuch des abgelaufenen Geschäftsjahres auch zu einer Erfassung der entsprechenden Kosten der umgesetzten Produkte geführt haben. (3 Stunden).
5. Vergleiche (je Produktgruppe) den prozentualen Anteil der Material- und Fertigungseinzelkosten sowie der produktionsbezogenen Gemeinkosten am Endbestand der Vorräte mit den entsprechenden Anteilen an den Gesamtkosten der umgesetzten Produkte (3 Stunden).
6. Überprüfe die korrekte Periodenabgrenzung durch die Untersuchung der Transaktionen um den Inventurstichtag herum und berücksichtige hierzu die Aufzeichnungen aus der Inventurbeobachtung (4 Stunden).
7. Untersuche, ob die Rückläufe und Gutschriften (insbesondere diejenigen mit hohen Beträgen) der ersten vier Wochen des neuen Geschäftsjahres auf Transaktionen beruhen, die dem alten Geschäftsjahr zuzuordnen sind (3 Stunden).
8. Führe eine gründliche Recherche in einschlägigen Fachzeitschriften und Wirtschaftsmagazinen durch, um festzustellen, inwieweit die gesamte Branche für Fertighäuser von neueren Entwicklungen betroffen ist (3 Stunden).
9. Vergleiche die systemseitig hinterlegten Plankosten des aktuellen Geschäftsjahres mit denen des Vorjahres (2 Stunden).
10. Teste die Bewertung der Rohstoffe unter Bezugnahme auf die laufende Bestandskartei und Lieferantenrechnungen (3 Stunden).
11. Vergleiche die Absatzstruktur sowie die Verkaufspreise pro Stück je Produktvariante mit dem Vorjahr (2,5 Stunden).
12. Stelle die Gewinnmargen nach Produktgruppen im Vergleich zum Vorjahr gegenüber (2,5 Stunden).
13. Vergleiche die Konten der Plankostenabweichung mit dem Vorjahr und kontrolliere die Verbuchung für das Geschäftsjahr 2005 (2 Stunden).
14. Verfolge schrittweise den Ablauf von den Versanddokumenten zu den Verkaufsrechnungen und den Verbuchungen in das Warenausgangsbuch, um sicherzustellen, dass alle Verkaufstransaktionen korrekt erfasst wurden (4 Stunden).
15. Untersuche Lieferantenrechnungen und stelle fest, ob Waren, die zum 31.12. noch nicht am Zielort der Liefervereinbarungen waren, fälschlicherweise in die Vorräte des abgelaufenen Geschäftsjahres gebucht wurden (4 Stunden).
16. Überprüfe die Vorräte oder die laufende Bestandskartei durch Testaufnahmen und Vermerkung der beschädigten, obsoleten oder überbevorrateten Posten (4 Stunden).
17. Vergleiche die systemseitig hinterlegten Plankosten mit den tatsächlichen Kosten, einschließlich Preistests bei Rohstoffen und Berechnung der tatsächlich angefallenen Fertigungs- und produktionsbezogenen Gemeinkosten (3 Stunden).
18. Führe ein Gespräch mit dem Leiter der Marketingabteilung und diskutiere mit ihm die aktuelle Marktsituation und die herangezogenen Strategien, um darauf zu reagieren (2,5 Stunden).
19. Untersuche die Aufzeichnungen aus der Inventurbeobachtung auf absolute Änderungen in den Beständen im Vergleich zum Vorjahr (3 Stunden).

## **Anhang 4: Grundversion einer Fallstudie**



# **Technische Universität Darmstadt**

**Fachgebiet für Rechnungswesen,  
Controlling und Wirtschaftsprüfung**

**Prof. Dr. Reiner Quick  
Dipl.-Kfm. Engin Kayadelen**

## **Fallstudie**

### **Durchführung von analytischen Prüfungen im Rahmen der Jahresabschlussprüfung**

**Bearbeitung bitte bis zum 18. April 2006**

**Kontaktadresse:**

Dipl.-Kfm. Engin Kayadelen  
Technische Universität Darmstadt  
Institut für Betriebswirtschaftslehre  
BWL 4  
Hochschulstr. 1

64289 Darmstadt

Telefon: (06151) 16-3523

E-Mail: [kayadelen@bwl.tu-darmstadt.de](mailto:kayadelen@bwl.tu-darmstadt.de)



## Vorwort

Sehr geehrte Damen und Herren,

herzlichen Dank, dass Sie sich die Zeit nehmen, diese Fallstudie zu bearbeiten. Bitte lesen Sie sich zunächst aufmerksam die vorliegenden Fallstudienunterlagen durch. Um sich anschließend online für die Fallstudienbearbeitung zu autorisieren, bitte ich Sie, bei Aufforderung folgendes zufällig generiertes Passwort zu verwenden (bitte berücksichtigen Sie hierbei auch Groß- und Kleinschreibung):

xxx

Bevor Sie jedoch mit der Bearbeitung beginnen, möchte ich Sie noch darauf hinweisen, dass diese Untersuchung ausschließlich der Durchführung einer empirischen Studie als Bestandteil meiner Dissertationsschrift dient. Es handelt sich hierbei insbesondere nicht um eine interne (Schulungs-) Maßnahme Ihrer Prüfungsgesellschaft. Obwohl sich die verwendeten Begrifflichkeiten und Vorgehensweisen – soweit möglich – an der Prüfungspraxis orientieren, ist es möglich, dass sich diese von internen Eigenheiten differenzieren.

Weiterhin bitte ich Sie, die Fallstudienbearbeitung alleine vorzunehmen und keine Informationen hinsichtlich Ihrer Vorgehensweise oder einer möglichen Lösung mit anderen Mitarbeitern Ihrer Prüfungsgesellschaft auszutauschen, da diese möglicherweise ebenfalls an der Studie teilnehmen.

Nochmals herzlichen Dank für Ihre Unterstützung. Ich werde Ihnen – sofern Sie dies wünschen – schnellstmöglich die Ergebnisse der Untersuchung zukommen lassen.

Engin Kayadelen

Darmstadt, im Februar 2006

## 1) Hintergrundinformationen

### Informationen zur Jahresabschlussprüfung der *Homelywood AG*:

Gehen Sie davon aus, dass Sie als Abschlussprüfer der *Homelywood AG* für das Geschäftsjahr 2005 vorgehen sind und das Unternehmen bereits im dritten Geschäftsjahr prüfen. Hierbei wenden Sie einen risiko-orientierten Prüfungsansatz an.

Während der letzten beiden Jahresabschlussprüfungen hat sich zwischen Ihnen als Abschlussprüfer und dem kaufmännischen Geschäftsführer des Mandanten ein gutes Arbeitsverhältnis entwickelt. Die Mitarbeiter des Mandanten arbeiten zuverlässig und kompetent.

Im Rahmen der aussagebezogenen Prüfungen wurden in den Geschäftsjahren 2003 und 2004 keine wesentlichen Fehler aufgedeckt, so dass Sie die vom Mandanten vorgelegten Jahresabschlüsse 2003 bzw. 2004 ohne wesentliche Nachbuchungen mit einem uneingeschränkten Bestätigungsvermerk testierten. Das interne Kontrollsystem wurde in den Geschäftsjahren 2003 und 2004 umfassend geprüft und dokumentiert. Insgesamt können die internen Kontrollen als ausreichend und wirksam beschrieben werden.

### Informationen aus der Aufnahme der wesentlichen Prozesse:

Die *Homelywood AG* ist ein Produzent für Außenwände von Fertighäusern und vertreibt ihre Produkte ausschließlich in der EU. Bisher produzierte das Unternehmen Außenwände in drei unterschiedlichen Stärken: Produkt A weist eine Stärke von 25 cm, Produkt B von 30 cm und Produkt C von 40 cm auf. Die Höhe aller drei Produkte beläuft sich auf 5 Meter. Die Breite kann entweder 4 Meter (Variante 1) oder 6 Meter (Variante 2) betragen. Das bisherige Produktsortiment gestaltete sich demnach entsprechend nachfolgender Tabelle:

| Produktnummer | Produkt | Variante | Maße (Höhe×Breite×Dicke) in Meter |
|---------------|---------|----------|-----------------------------------|
| 1             | A       | 1        | 5,00×4,00×0,25                    |
| 2             | A       | 2        | 5,00×6,00×0,25                    |
| 3             | B       | 1        | 5,00×4,00×0,30                    |
| 4             | B       | 2        | 5,00×6,00×0,30                    |
| 5             | C       | 1        | 5,00×4,00×0,40                    |
| 6             | C       | 2        | 5,00×6,00×0,40                    |

Die Außenwände bestehen aus einer doppelten Gipskartonplatte und einem Innenkern aus Kunststoff, der hauptsächlich der Geräusch- und Wärmedämmung dient. Im Rahmen des dreistufigen Produktionsprozesses werden in einem ersten Schritt zunächst die Gipskartonplatten auf die notwendigen Maße zugeschnitten. Im zweiten Schritt werden jeweils zwei Gipskartonplatten miteinander verbunden. Im abschließenden dritten Schritt wird der Innenkern in den Hohlraum zwischen den Gipskartonplatten gespritzt. Anschließend werden die fertigen Produkte in das Lagerhaus transportiert.

Die Produktionsstätte befindet sich in Weiterstadt und weist eine Größe von 25.000 m<sup>2</sup> auf. Hier befinden sich auch die Räumlichkeiten der Geschäftsführung, Verwaltung und Vertrieb. Die Lagerhallen befinden sich ebenfalls in unmittelbarer Nähe und umfassen zusätzlich 15.000 m<sup>2</sup>. Insgesamt beschäftigt die *Homelywood AG* rund 1.100 Mitarbeiter in Weiterstadt.

### Informationen zu dem Finanzbuchführungssystem und der Ermittlung der Herstellungskosten:

Zur Bewertung der Vorratsbestände sowie der Kosten der umgesetzten Einheiten (Umsatzkosten) werden im Finanzbuchführungssystem für jede Produktausführung Plankosten hinterlegt. Entsprechend den unternehmensinternen Vorgaben zur Ermittlung der Herstellungskosten für den handelsrechtlichen Jahresabschluss umfassen die Plankosten die Material- und Fertigungseinzelkosten und die produktionsbezogenen Gemeinkosten. Zum Ende eines jeden Geschäftsjahres werden die Plankosten vom Controller auf ihre Richtigkeit überprüft. Differenzen zwischen Plankosten und den tatsächlichen Kosten werden auf das Konto Plankostenabweichung gebucht und anteilmäßig auf die Vorratsbestände sowie Umsatzkosten verrechnet.

### Geschehnisse im abgelaufenen Geschäftsjahr:

Seit Geschäftsjahresbeginn kommen bei der *Homelywood AG* neue Kunststoffe für den Innenkern der Wände zum Einsatz, welche sich durch höhere Dämmeigenschaften auszeichnen. Trotz höherer Dämmeigenschaften kann der Kunststoff aufgrund von Verhandlungsgeschick zum gleichen Preis wie der bisherige Kunststoff beschafft werden. Aufgrund vergleichbarer Verarbeitungseigenschaften konnte der Produktionsprozess ohne wesentliche Änderung beibehalten werden. Die *Homelywood AG* hat über diese Änderung hinaus keine weiteren Modifikationen erfahren. Insbesondere wurden keine strukturellen und personellen Änderungen gegenüber dem Vorjahr vorgenommen.

## 2) Jahresabschlussdaten der Homelywood AG nach HGB

### Bilanz

| <b>Aktiva</b><br><b>in Euro</b>                      | <b>2005</b><br><b>(ungeprüft)</b> | <b>2004</b>       | <b>Abw.</b><br><b>%</b> |
|------------------------------------------------------|-----------------------------------|-------------------|-------------------------|
| Immaterielle Vermögensgegenstände                    | 3.364.147                         | 3.628.405         | -7%                     |
| Sachanlagen                                          | 13.344.486                        | 13.674.517        | -2%                     |
| Finanzanlagen                                        | 1.283.493                         | 1.084.862         | 18%                     |
| <b>Anlagevermögen</b>                                | <b>17.992.126</b>                 | <b>18.387.784</b> | <b>-2%</b>              |
| Vorräte                                              | 4.452.607                         | 4.167.514         | 7%                      |
| Forderungen aus Lieferungen und Leistungen           | 6.226.055                         | 5.006.266         | 24%                     |
| Übrige Forderungen und Sonstige Vermögensgegenstände | 995.812                           | 813.338           | 22%                     |
| Wertpapiere                                          | 206.840                           | 168.336           | 23%                     |
| Flüssige Mittel                                      | 1.711.666                         | 1.125.296         | 52%                     |
| <b>Umlaufvermögen</b>                                | <b>13.592.979</b>                 | <b>11.280.751</b> | <b>20%</b>              |
| <b>Latente Steuern</b>                               | <b>983.917</b>                    | <b>958.060</b>    | <b>3%</b>               |
| <b>Rechnungsabgrenzungsposten</b>                    | <b>191.083</b>                    | <b>183.367</b>    | <b>4%</b>               |
| <b>Summe Aktiva</b>                                  | <b>32.760.105</b>                 | <b>30.809.962</b> | <b>6%</b>               |

| <b>Passiva</b><br><b>in Euro</b>                        | <b>2005</b><br><b>(ungeprüft)</b> | <b>2004</b>       | <b>Abw.</b><br><b>%</b> |
|---------------------------------------------------------|-----------------------------------|-------------------|-------------------------|
| Gezeichnetes Kapital                                    | 1.430.838                         | 1.430.838         | 0%                      |
| Kapitalrücklage                                         | 4.903.381                         | 4.903.381         | 0%                      |
| Gewinnrücklagen                                         | 4.346.739                         | 3.573.406         | 22%                     |
| Bilanzgewinn                                            | 4.151.943                         | 3.081.217         | 35%                     |
| <b>Eigenkapital</b>                                     | <b>14.832.901</b>                 | <b>12.988.842</b> | <b>14%</b>              |
| Rückstellung für Pensionen und ähnliche Verpflichtungen | 6.308.190                         | 6.129.212         | 3%                      |
| Übrige Rückstellungen                                   | 2.348.092                         | 2.382.516         | -1%                     |
| <b>Rückstellungen</b>                                   | <b>8.656.282</b>                  | <b>8.511.728</b>  | <b>2%</b>               |
| Finanzverbindlichkeiten                                 | 3.251.513                         | 3.389.866         | -4%                     |
| Verbindlichkeiten aus Lieferungen und Leistungen        | 3.070.464                         | 3.007.231         | 2%                      |
| Übrige Verbindlichkeiten                                | 2.657.028                         | 2.607.520         | 2%                      |
| <b>Verbindlichkeiten</b>                                | <b>8.979.005</b>                  | <b>9.004.616</b>  | <b>0%</b>               |
| <b>Latente Steuern</b>                                  | <b>142.337</b>                    | <b>152.183</b>    | <b>-6%</b>              |
| <b>Rechnungsabgrenzungsposten</b>                       | <b>149.581</b>                    | <b>152.593</b>    | <b>-2%</b>              |
| <b>Summe Passiva</b>                                    | <b>32.760.105</b>                 | <b>30.809.962</b> | <b>6%</b>               |

### Gewinn- und Verlustrechnung

| <b>in Euro</b>                                         | <b>2005</b><br><b>(ungeprüft)</b> | <b>2004</b>       | <b>Abw.</b><br><b>%</b> |
|--------------------------------------------------------|-----------------------------------|-------------------|-------------------------|
| Umsatzerlöse                                           | 38.741.643                        | 35.577.998        | 9%                      |
| Kosten der umgesetzten Produkte (Umsatzkosten)         | -27.098.451                       | -25.412.167       | 7%                      |
| <b>Bruttoergebnis vom Umsatz (Rohgewinn)</b>           | <b>11.643.193</b>                 | <b>10.165.830</b> | <b>15%</b>              |
| Vertriebskosten                                        | -3.364.178                        | -3.180.666        | 6%                      |
| Allgemeine Verwaltungskosten                           | -2.405.862                        | -2.291.635        | 5%                      |
| Sonstige betriebliche Erträge                          | 332.507                           | 324.191           | 3%                      |
| Sonstige betriebliche Aufwendungen                     | -476.672                          | -401.805          | 19%                     |
| Beteiligungsergebnis                                   | 837                               | 921               | -9%                     |
| <b>Betriebliches Ergebnis</b>                          | <b>5.729.825</b>                  | <b>4.616.836</b>  | <b>24%</b>              |
| <b>Finanzergebnis</b>                                  | <b>-174.184</b>                   | <b>-148.056</b>   | <b>18%</b>              |
| <b>Ergebnis gewöhnlicher Geschäftstätigkeit (EBIT)</b> | <b>5.555.641</b>                  | <b>4.468.779</b>  | <b>24%</b>              |
| Steuern vom Einkommen und vom Ertrag                   | -630.365                          | -550.896          | 14%                     |
| <b>Jahresüberschuss</b>                                | <b>4.925.276</b>                  | <b>3.917.884</b>  | <b>26%</b>              |
| Einstellung in Gewinnrücklagen                         | -773.333                          | -836.667          | -8%                     |
| <b>Bilanzgewinn</b>                                    | <b>4.151.943</b>                  | <b>3.081.217</b>  | <b>35%</b>              |

Anhand dieser Jahresabschlussdaten ermittelt der Senior Manager eine Wesentlichkeitsgrenze von 387.416 € (1% der Umsatzerlöse des abgelaufenen Geschäftsjahres) sowie folgende jahresabschlussbasierte Kennzahlen:

|                                            |                                                                            | 2005<br>(ungeprüft) | 2004<br>(geprüft) |
|--------------------------------------------|----------------------------------------------------------------------------|---------------------|-------------------|
| Liquidität 3. Grades =                     | $\frac{\text{Umlaufvermögen}}{\text{kurzfristige Verb.}}$                  | 2,27                | 1,88              |
| Liquidität 2. Grades =                     | $\frac{\text{Umlaufvermögen} - \text{Vorräte}}{\text{kurzfristige Verb.}}$ | 1,53                | 1,18              |
| Umschlagshäufigkeit der Vorräte =          | $\frac{\text{Umsatzkosten}}{\text{Endbestand Vorräte}}$                    | 6,09                | 6,10              |
| Umschlagshäufigkeit der Forderungen aLuL = | $\frac{\text{Umsatzerlöse}}{\text{Endbestand Forderungen aLuL}}$           | 6,22                | 7,11              |
| Rohgewinnspanne =                          | $\frac{\text{Bruttoergebnis vom Umsatz (Rohgewinn)}}{\text{Umsatzerlöse}}$ | 30,05%              | 28,57%            |
| Umsatzrentabilität =                       | $\frac{\text{EBIT}}{\text{Umsatzerlöse}}$                                  | 14,34%              | 12,56%            |

### 3) Hinweise zur Prüfung des abgelaufenen Geschäftsjahres 2005

Zur Durchführung von analytischen Prüfungshandlungen erhalten Sie von Ihrem Senior Manager folgende Sollwerte für das zu prüfende Geschäftsjahr. Zu Vergleichszwecken werden diese im Folgenden den entsprechenden Werten des zu prüfenden Geschäftsjahres gegenübergestellt:

|                                            |                                                                            | Sollwert | 2005<br>(ungeprüft) |
|--------------------------------------------|----------------------------------------------------------------------------|----------|---------------------|
| Liquidität 3. Grades =                     | $\frac{\text{Umlaufvermögen}}{\text{kurzfristige Verb.}}$                  | 2,14     | 2,27                |
| Liquidität 2. Grades =                     | $\frac{\text{Umlaufvermögen} - \text{Vorräte}}{\text{kurzfristige Verb.}}$ | 1,53     | 1,53                |
| Umschlagshäufigkeit der Vorräte =          | $\frac{\text{Umsatzkosten}}{\text{Endbestand Vorräte}}$                    | 7,56     | 6,09                |
| Umschlagshäufigkeit der Forderungen aLuL = | $\frac{\text{Umsatzerlöse}}{\text{Endbestand Forderungen aLuL}}$           | 6,22     | 6,22                |
| Rohgewinnspanne =                          | $\frac{\text{Bruttoergebnis vom Umsatz (Rohgewinn)}}{\text{Umsatzerlöse}}$ | 28,07%   | 30,05%              |
| Umsatzrentabilität =                       | $\frac{\text{EBIT}}{\text{Umsatzerlöse}}$                                  | 12,36%   | 14,34%              |

Sie können davon ausgehen, dass der Senior Manager bei der Ermittlung der Sollwerte alle wesentlichen Informationen, wie die Daten der untestierten Quartalsabschlüsse I-III des abgelaufenen Geschäftsjahres, Daten des Vorjahresabschlusses und Brancheninformationen, berücksichtigt hat, und dass die Sollwerte zuverlässig sind.

Der Senior Manager ist der Auffassung, dass die Abweichung in der Rohgewinnspanne als unerwartet und wesentlich zu bewerten ist, und fordert, dass diese Abweichung untersucht werden muss.

In einem Gespräch mit dem Geschäftsführer der *Homelywood AG* weisen Sie auf die unerwartete Abweichung in der Rohgewinnspanne hin. Der Geschäftsführer erklärt Ihnen, dass

„...die steigenden Preise für Heizöl und Gas sowie der allgemeine Trend zur Begrenzung der CO<sub>2</sub>-Emission europaweit die Nachfrage nach energieeffizienten Fertighäusern erhöht haben. Seit Verwendung der neuen, energieeffizienten Kunststoffe konnte das Unternehmen von dieser gestiegenen Nachfrage profitieren und die Gewinnmargen erhöhen.“

## Bearbeitung der Fallstudie

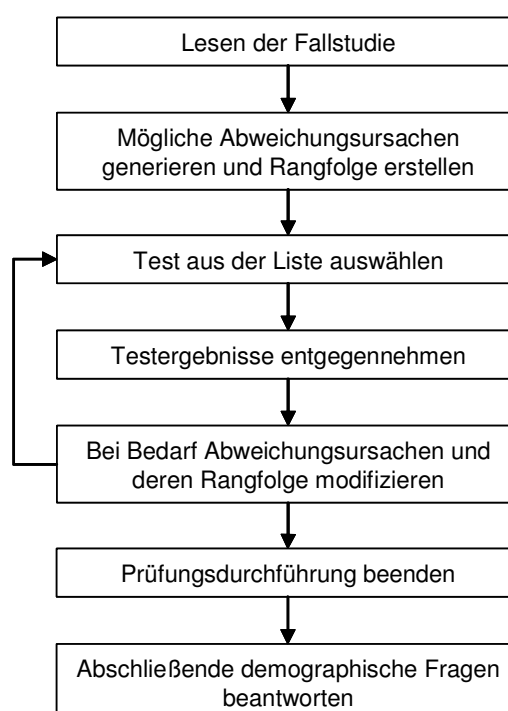
### 1) Aufgabenstellung

Ihre Aufgabe besteht darin, den Grund für die festgestellte Abweichung in der Rohgewinnspanne zu identifizieren. Gehen Sie dabei davon aus, dass 90% der festgestellten Abweichung durch einen einzelnen Grund oder das mehrfache Auftreten des gleichen Grundes hervorgerufen wurde. Bei diesem Grund kann es sich sowohl um einen Fehler im Rechnungswesen handeln als auch um Faktoren, die auf Änderungen in der Unternehmensumwelt zurückzuführen sind (sog. Nicht-Fehler).

Um den Grund für die Abweichung festzustellen, wird Ihnen im Rahmen der *internetbasierten Aufgabenbearbeitung* eine Liste mit 19 möglichen Tests zur Verfügung gestellt. Aus dieser Liste können Sie eine bestimmte Anzahl von Tests wählen. Anschließend werden Sie über die entsprechenden Testergebnisse in Kenntnis gesetzt. Jeder Test nimmt einen bestimmten zeitlichen Aufwand in Anspruch, der ebenfalls der Liste zu entnehmen ist. Insgesamt steht Ihnen ein Zeitbudget von 18 Stunden zur Verfügung, welches nach Möglichkeit um nicht mehr als 10% überschritten werden sollte.

### 2) Internetbasierte Aufgabenbearbeitung

Nachdem Sie sich online für die Bearbeitung der Fallstudie autorisiert haben, führt Sie das System durch die einzelnen notwendigen Schritte. Der Ablauf der Fallstudienbearbeitung wird Ihnen anhand der folgenden Abbildung grafisch dargestellt:



# Lebenslauf

## Engin Kayadelen

\*2. September 1973 in Gelsenkirchen

### Schul Ausbildung

08/1980 - 06/1984

Grundschule Tigelschule in Essen

08/1984 - 06/1994

Gymnasium Essen Nord-Ost

### Hochschulausbildung

10/1994 - 09/1995

Beginn des Studienfaches Wirtschaftsinformatik mit anschließendem Wechsel zum Studiengang Betriebswirtschaftslehre

10/1995 - 10/2001

Studium der Betriebswirtschaftslehre an der Universität Essen

#### *Abschluss*

Diplom-Kaufmann, Gesamtnote: 1,7

#### *Studienschwerpunkte*

Wirtschaftsprüfungswesen, Unternehmensrechnung & Controlling, Wirtschaftsrecht

#### *Diplomarbeit*

"Emissionsprospekte am Neuen Markt - Unter besonderer Berücksichtigung der Prognosepublizität" (ausgezeichnet mit dem Preis der Genossenschaftsbank Essen eG für Wirtschaftswissenschaften)

### Berufliche Tätigkeiten

01/2002 - 09/2002

Prüfungsassistent bei der Deloitte & Touche GmbH in Düsseldorf

10/2002 - 09/2003

Wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschaftsprüfungswesen der Universität Duisburg-Essen

10/2003 - 09/2004

Wissenschaftlicher Mitarbeiter am Institut für Revisionswesen der Westfälischen Wilhelms-Universität Münster

04/2004 - 09/2004

Dozent an der Verwaltungs- und Wirtschaftsakademie (VWA) Münster

10/2004 - 06/2007

Wissenschaftlicher Mitarbeiter am Fachgebiet für Rechnungswesen, Controlling und Wirtschaftsprüfung der Technischen Universität Darmstadt

### Promotion

06/2007

Abgabe der Dissertationsschrift

Titel: Zur Durchführung von analytischen Prüfungshandlungen im Rahmen der Jahresabschlussprüfung

12/2007

Disputation

## **Veröffentlichungen**

Kennzahlengestütztes Value Added Reporting in der Geschäftsberichts-  
publizität der Eurostoxx-50-Unternehmen. Die Wirtschaftsprüfung 2008,  
S. 156-163 (zusammen mit Reiner Quick und Martin Flashaar-Bloedorn).

Zur Aussagefähigkeit von Prognosen in Emissionsprospekten am Neuen  
Markt. Die Wirtschaftsprüfung 2002, S. 949-965 (zusammen mit Reiner  
Quick).

Analytische Prüfungen, in: K.-U. Marten, R. Quick, K. Ruhnke (Hg.): Lexi-  
kon der Wirtschaftsprüfung, Stuttgart 2006, S. 25-31.

Anhang, Prüfung, in: K.-U. Marten, R. Quick, K. Ruhnke (Hg.): Lexikon  
der Wirtschaftsprüfung, Stuttgart 2006, S. 37-43.

Eigenkapitalveränderungsrechnung, Prüfung, in: K.-U. Marten, R. Quick,  
K. Ruhnke (Hg.): Lexikon der Wirtschaftsprüfung, Stuttgart 2006, S. 220-  
224.

Haftungsverhältnisse und sonstige finanzielle Verpflichtungen, Prüfung,  
in: K.-U. Marten, R. Quick, K. Ruhnke (Hg.): Lexikon der Wirtschaftsprü-  
fung, Stuttgart 2006, S. 368-372.

Prognoseprüfung, in: K.-U. Marten, R. Quick, K. Ruhnke (Hg.): Lexikon  
der Wirtschaftsprüfung, Stuttgart 2006, S. 571-575.

Vermögenslage, Prüfung, in: K.-U. Marten, R. Quick, K. Ruhnke (Hg.):  
Lexikon der Wirtschaftsprüfung, Stuttgart 2006, S. 842-846.

## **Mitwirkung an den folgenden Buchprojekten:**

Wirtschaftsprüfung – Grundlagen des betriebswirtschaftlichen Prüfungs-  
wesens nach nationalen und internationalen Normen, K.-U. Marten, R.  
Quick, K. Ruhnke (Hg.), zweite und dritte Auflage.

Lexikon der Wirtschaftsprüfung – nach nationalen und internationalen  
Normen, K.-U. Marten, R. Quick, K. Ruhnke (Hg.).

Bilanzierung in Fällen – Grundlagen, Aufgaben und Lösungen, R. Quick  
(Hg.), zweite und dritte Auflage.

Doppelte Buchführung – Grundlagen, Übungsaufgaben, Lösungen, R.  
Quick, H.-J. Wurl (Hg.).